USING DATA MINING IN FUND RAISING

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

ABSTRACT

The objective of this study is to apply data mining techniques to fund raising by utilizing data-mining and analyzing software to draw correlations between data variables, and thus mine for information. With the advances in technology and the rapidly growing economy, fund raising has gone professional. There are national associations for fund raisers, providing training and refresher courses. Fund raising consultants and IT companies compete to market their techniques and software. Data mining tools are software tools used to query information in a database or data warehouse. Data mining tools allow the manipulation of data or information to support decision making tasks. By applying data mining techniques to fund raising, we try to find a pattern in the donor database, and target the potential donors.

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 2000, the American Association of Fundraising Counsel reported that charitable giving in the United States surpassed $190 billion as the 20th century came to a close. Corporations gave more than $11 billion annually to worthy causes; foundations gave more than $19 billion; and individuals gave by far the largest amount at $143 billion, including an additional $15 billion in bequests. Of the total amount contributed, more than $27 billion was given to public and private education, which ranks second only to religious organizations as a recipient of grants and gifts.

Tools for Fund Raising

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 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 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.

Methodology

In this study, data mining techniques were applied to fund raising by using PolyAnalyst a data mining and analyzing software that has not specifically been designed for fund raisers. Data mining represents a new and promising branch of Artificial Intelligence (AI) that embraces technologies aimed at the automated extraction of knowledge (meaningful patterns and rules) from databases. PolyAnalyst is a multi-strategy data mining system that implements a broad variety of mutually complementing methods for the automatic data analysis. PolyAnalyst works with data extracted from flat files or relational databases. The data used for this study was obtained from a four year college foundation in Montana. A database of University alumni, with demographic information such as age, city, state, marital status, monetary contribution and year of graduation was obtained. The data also contained contact information such as addresses, telephone numbers and e-mail addresses. The aim of the study is to try and find a pattern in the donor database, to help target potential donors. 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. This data was then categorized using Microsoft excel.

Analysis

PolyAnalyst automatically finds dependencies and laws hidden in data, presenting them explicitly in the form of rules and algorithms. The Find Dependencies algorithm was used to find relationships and dependencies in the data and select influencing (independent) variables and a dependent variable. The dependent variable selected for the test was the donor contribution. The test also allowed us to choose an algorithm type as strict or liberal. The “strict” rule was for finding which attributes are most important to the exploration and the “liberal” rule was for finding outlying points, errors, or exceptions to a rule. This study utilized the strict rule and the variables that were selected by the test as being most important as dependencies are as follows:

Most influencing parameters:	“Age”, "Marital Status", “Years since graduation”
P-value:	3.195e-060
N of points obeying the dependence:	1876
Std. deviation (total):	3982.965820 (107.39%)
Std. deviation (for FD-subpopulation):	414.946564 (55.86%)

70% of data points lay inside +-302.195618 interval of errors
90% of data points lay inside +-1193.978760 interval of errors
97% of data points lay inside +-7060.668457 interval of errors

Utilizing the variables found in the dependencies test, as the most important dependencies, a DECISION TREE test was performed. A decision tree (DT) is often used to represent a large family of machine learning algorithms for the automated construction of tree-like classification rules for categorizing structured data. DT works best for solving classification tasks. But is far less efficient to the discovery of numerical dependencies. The result obtained on running the test was as follows:

Number of non-terminal nodes: 1
Number of leaves: 2
Depth of constructed tree: 1

The report obtained made no predictive sense and was discarded as invalid. The next test that was run over the entire data was a Neural Network, called the PolyNet Predictor. PolyNet Predictor generates a network of nodes, each of which contains a mathematical expression, which can be used to predict the value of one attribute based on the values of several others. The PolyNet Predictor report gives a variety of significance and accuracy measures, as well as a few statistics on the neural network itself. The report begins with the Significance Index, Standard Error, R-squared, and Standard Deviation for each rule. These are all measures of the accuracy and significance of the rules produced. The process included attributes such as gender, age, marital status, the class, years since graduation, the major, city, and state. The exploration created mining model with the following exploration parameters:

Critical F-Ratio 1
Minimum part of predicted values, % 50
Order 3

Significance index:	291.2
Standard error:	0.9278
R-squared:	0.1392
Standard deviation:	0.4606
Points processed:	5979
Number of network layers:	1
Number of network nodes:	3

The results obtained for the Significance Index, Standard Error, R-squared, and Standard Deviation did not indicate an accurate prediction. The output was also obtained in the form of a gain chart, a prediction vs. real chart, a residual chart and a prediction and real vs. counter chart, which did not indicate good predictions. The last test was the Nearest Neighbor, a memory-based classification system, which assigns values to data points based on their “proximity” to other data points. This test did not provide good results.

Conclusions and Future Research

PolyAnalyst was found to be powerful analytical software. It allowed us to create charts and objects at the touch of a button, which revealed relationships between any two or three variables. The study’s failure to provide good predictive results could be attributed to using just demographic information and lack of more information regarding the donors, which has specific correlation to contribution. Utilizing a more powerful data mining software or a data mining software specifically designed for fund raising may result in predicting more accurate relationships, which could be the subject of a future research. References are available upon request.