Data Mining Framework for Direct Marketing: A Case Study of Bank Marketing

Lilian Sing'oei[1] and Jiayang Wang[2]

[1] School of Information Science and Engineering, Central South University
Changsha, 410083, China

[2] School of Information Science and Engineering, Central South University
Changsha, 410083, China

Abstract

Direct marketing has generated an increasing interest among academics and practitioners over the past few years. This is due to competitive market environment, advancement in technology and changing behavior of customers which are difficult to predict. Despite of numerous studies that have provided important insights into the direct marketing, the understanding of this topic of growing interest and importance still remains deficient. Therefore, the objective of this paper is to provide a comprehensive framework to guide research efforts focusing on direct marketing strategy and aid practitioners in their quest to achieve direct marketing success using data mining methods. The framework builds on the literature from direct marketing concepts and data mining methods that provides a systematic approach to users who have little knowledge in data mining in order to carry out effective marketing campaigns. A case study on bank marketing campaigns was used for evaluating the feasibility of the framework.

Keywords: Data Mining, Direct marketing, Decision Tree, Data Mining Framework.

1. Introduction

Direct marketing is the process of identifying potential buyers of certain products and promoting the products accordingly [15]. There are two methods of advertisement and promotions, namely mass marketing and direct marketing. Mass marketing targets general public by using media such as newspapers, radios and television. This results in high waste and low response rate from customers who will actually buy the product [14]. In today's competitive market place, mass marketing is no longer efficient and reliable method. Hence, marketers are shifting their focus from traditional mass marketing to direct marketing. Direct marketing studies customers' characteristics, needs and selects certain customers as their target for promotion. To counteract the rising costs and declining response rates from customers, direct marketing use predictive modeling by analyzing the customers data
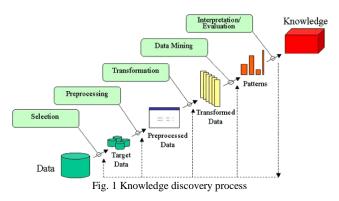
(demographic and historical purchase data) in order to select the customers who are more likely to respond to a promotion [8], [9], [10] which give higher response rate and it is an effective method for marketing.

Large amounts of data are generated every day in many organizations. These data can be used to establish and maintain direct relationship with the customers in order to target them individually for specific offers. To extract hidden predictive information from large amounts of data, Organizations are now realizing the importance of data mining in their strategic planning and successful application of data mining methods that can generate useful knowledge to organizations[6], [11]. There are a lot of research papers on computational and theoretical aspects of direct marketing though little efforts have been put on technological aspects of applying data mining in the direct market process. This is due to the complexity of the data mining models that make it difficult for marketers to use and understand them [7]. In addition, the marketers have little knowledge on data mining skills; hence there is a need to develop a simplified framework to guide marketers in making use of data mining methods for direct marketing.

The rest of paper is organized as follows: section 2 provides an overview of data mining. In section 3, we explain classification in data mining, in section4; we describe direct marketing framework using data mining and highlight the main phases and functions involved in each phase. In section 5, we present a case study of direct marketing in banks. In section 6, we provide experimental results and finally, we present a conclusion and references.

2. Data Mining

Data mining is the use of sophisticated data analysis tools to discover previously unknown, valid patterns and relationships in large data sets. These tools include mathematical algorithms, statistical models and machine learning methods. The growing interest of data mining in

business is enhanced by number of technological factors. The growing amounts of customer data are collected and made accessible in data repositories and data marts. Powerful new data analysis algorithms are discovered by researchers from statistical pattern recognition and artificial intelligence field such as machine learning, neural networks and evolutionary computation. Recently, ordinary office computers are powerful enough to run these advanced data mining algorithms [5], [6]. Data mining tools takes data and construct a model as a representation of the reality. The aim of data mining is to solve a business problem [9]. In direct marketing, data mining are used to reveal a specific class of customers which are most likely to be interested in a particular product. This will allow the planning of a direct marketing campaign aimed toward a specific class of customers with the aim of achieving higher response [7]. The data mining process is sometimes referred to as Knowledge Discovery in Databases (KDD). The "KDD" refers to the overall process of discovering useful knowledge from data. Knowledge discovery in database is the process of identifying valid, novel, potentially useful and understandable patterns or models in data [x]. Data mining itself is a step in knowledge discovery process. The steps involved in knowledge discovery are:

Fig. 1 Knowledge discovery process

Data Selection: The data relevant to the analysis is decided and retrieved from the various data locations.

Data Preprocessing: This stage consists of:

Data Cleaning: this is the removing of noisy data and irrelevant data from the data collected.

Data Transformation: This is where the selected data is transformed into forms appropriate for the mining procedure.

Data Mining: It is the crucial step in which clever techniques are applied to extract potentially useful patterns. The decision is made about the data mining technique to be used.

Interpretation and Evaluation: In this step, interesting patterns representing knowledge are identified based on given measures. The discovered knowledge is visually presented to the user. This essential step uses visualization techniques to help users understand.

3. Classification in Data Mining

Classification is the most commonly applied data mining technique, which employs a set of pre-classified examples to develop a model that can classify the population of records at large [5]. Classification is a data mining function that assigns items in a collection to target categories or classes. The goal of classification is to accurately predict the target class for each case in the data. Basically classification is used to classify each item in a set of data into one of predefined set of classes or groups. Classification method makes use of mathematical techniques such as decision trees, linear programming, neural network and statistics [5]. A classification task begins with a data set in which the class assignments are known. In training process, a classification algorithm finds relationships between the values of the predictors and the values of the target. Different classification algorithms use different techniques for finding relationships. These relationships are summarized in a model, which can then be applied to a different data set in which the class assignments are unknown. Classification models are tested by comparing the predicted values to known target values in a set of test data. The historical data for a classification project is typically divided into two data sets: one for building the model; the other for testing the model.

3.1 Decision Tree

A decision tree is a popular classification technique that results in flowchart like tree structure where each node denotes test on an attribute value and each branch represents an outcome of test. The leaves represent classes. Using training data decision tree generates a tree that consists of nodes that are rules and each leaf node represents a classification or decision. The data usually plays important role in determining the quality of the decision tree. If there are number of classes, then there should be sufficient training data available that belongs to each of the classes. Decision trees are predictive models, used to graphically organize information about possible options, consequences and end value. They are used in for calculating probabilities.

3.2 C 4.5 Decision tree Algorithm

C4.5 builds a decision tree from set of training data using the concept of Information entropy. The training data is a

set S = s_1, s_2... of already classified samples. Each sample $s_1 = x_1, x_2$... is a vector where x_1, x_2... represent attributes or features of the sample. The training data is augmented with a vector C = c_1, c_2... where c_1, c_2... represent the class to which each sample belongs. At each node of the tree, C4.5 chooses one attribute of the data that most effectively splits its set of samples into subsets enriched in one class or the other. Its criterion is the normalized information gain (difference in entropy) that results from choosing an attribute for splitting the data. The attribute with the highest normalized information gain is chosen to make the decision. The C4.5 algorithm then recurses on the smaller sub lists. In general, steps in C4.5 algorithm to build decision tree are:

1. Choose attribute for root node
2. Create branch for each value of that attribute
3. Split cases according to branches
4. Repeat process for each branch until all cases in the branch have the same class.

4. Direct Marketing Framework Using Data Mining

The data mining framework for direct marketing consists of five phases outlined below:

Understanding the problem domain- The starting point of any data mining project is to understand the business goals and requirements of the business. The next step is to translate the data mining problem into a preliminary plan in order to attain the project objective. In this stage the problem is defined in details and selection of data mining tools to be used in the later process is identified. The four main data mining activities involved in this phase are trend analysis, customer profiling, project goals and deviation analysis [1], [3].

Data Preprocessing- Data preparation accounts for 60%-80% of the time spent on a data mining project. Data preprocessing consists of four tasks namely:
Data cleaning: Raw data is often incomplete, inconsistent and contain many errors. Data cleaning is achieved by filling in missing values, smoothing the noisy data and resolving the inconsistencies in the data therefore improving the quality of data.

Attribute or feature selection: In this stage one has to select a relevant attribute that appears to be a good predictor of the class and remove irrelevant or redundant attributes.

Data transformation: This is the transformation of raw data into an understandable format e.g. Microsoft excel, database etc. for further analysis and processing.
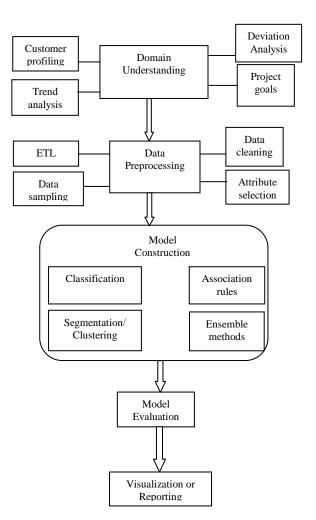

Fig.2 Data mining framework for direct marketing

ETL: this stands for Extract, Transform and Load. One has to extract data from the outside sources, transform it to fit into operational needs and load it into the end target e.g. data base or data warehouse.

Model construction- Once the historical data has been collected, the next thing is model construction. This is achieved by applying different models to the same data set and then comparing their performance to choose the best model. The main aim of creating a model is to predict the purchasing behavior of customers. Decision tree are the best and popular models for classification. This is because they have the ability to generate rules that can be translated into natural language. Complex decision trees can be dealt with by following the path through the tree to a particular leaf which makes the explanation of any particular classification or prediction relatively easy [5].

Model Evaluation- Model Evaluation is an important part of the model development process. It helps to find the best

model that represents our data and how well the chosen model will work in the future. There are two methods of evaluating models in data mining, Hold-Out and Cross-Validation.

Hold-out method is meant for large data sets and it is divided into three parts:
Training set- is a subset of the data set used to build predictive models.
Validation set- is a subset of data set used to assess the performance of the model built in the training phase. It provides a test platform for fine tuning model's parameters and selecting the best performance model. Not all modeling algorithms need a validation set.
Test set- a subset of the dataset to assess the likely future performance of a model. If a model fit to the training set much better than it fits the test set, over fitting is probably the cause.

Cross validation is used when only a limited amount of data is available; to achieve an unbiased estimate of the model performance we use *k*-fold cross-validation. In *k*-fold cross-validation, we divide the data into *k* subsets of equal size. We build models *k* times, each time leaving out one of the subsets from training and use it as the test set. If *k* equals the sample size, this is called "leave-one-out. The model with the best performance is selected for campaign [4].

Visualization of Marketing Knowledge-This is the application of a model for prediction or classification to new data. After a satisfactory model or set of models has been identified (trained) for a particular application, we usually want to apply those models so that predictions or predicted classifications can quickly be obtained for new data. We use lift chart to measure the effectiveness of a classification model which is calculated as the ratio between the results obtained with and without the model. Lift charts are visual aids for evaluating performance of classification models. The lift chart shows how much more likely we are to receive positive responses than if we contact a random sample of customers [1].

5. Case Study of Direct Marketing in Banks

The data set we used was from Portuguese banking institution and had 45,212 observations. Each observation represents a customer and is described by 17 attributes, both categorical and continuous. The variable "y" (target) describes whether the customer had subscribed to a term deposit. There were no missing values since detailed descriptions on the meaning of each variable were provided. All of the observations had received the direct

promotion. The classification goal is to predict if the client will subscribe to a term deposit (variable y).

All the experiments reported in this work were conducted using the IBM SPSS Modeler 14.2. Decision tree algorithm was chosen because they are powerful and popular tools for classification and prediction. The model constructed had great interoperability because rules can readily be expressed so that humans can understand them [2]. In addition, decision trees have the ability to generate rules that can be translated into natural language in contrast to neural networks. Complex decision trees can be dealt with by following the path through the tree to a particular leaf which makes the explanation of any particular classification or prediction relatively easy [2] [5].

The experiments were conducted using the supervised learning algorithm C5.0. The IBM SPSS modeler trained the model and finally passed on to us the validation results. We spent a lot of time on variable selection and variable transformation.

6. Experimental Results

We performed our experiments using 10-fold cross validation. The original dataset was randomly portioned into ten disjoint subsets. In each of the ten runs, nine of the subsets are combined to form the training set while the remaining subset forms the testing set. Based on the response rate, there are two classes, the positive class consists of the responders and negative class consists of non-responders. The highest responders are put into decile 1. The response rate is high showing that the supervised learning algorithm performs well using the training data being sampled. Lift is the most commonly used metric to measure the performance of targeting models in marketing applications. The target members selected are those likely to respond positively to a marketing offer. The results of the model are shown in figure 4.1 below. The x-axis represents the different percentiles x. The y-axis represents the response rate.
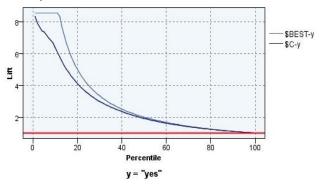

Fig. 3 Lift chart

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 2, No 2, March 2013
ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784
www.IJCSI.org

202

Fiq.4 Results for output field y

7. Conclusion

We demonstrated that data mining is an effective tool for direct marketing which can improve bank marketing campaigns. Most research papers focus on computational and theoretical aspects of direct marketing though little efforts have been put on technological aspects of applying data mining in the direct market process. The complexity of the data mining models makes it difficult for marketers to use it, hence; we outlined a simplified framework to guide marketers and managers in making use of data mining methods and focus their advertising and promotion on those categories of people in order to reduce time and costs. We explained all the steps and tasks that are carried out at each stage of the data mining framework. The case study we evaluated shows the practical use and usefulness of the model.

References

[1] M. J Shaw, C Subramaniam, G.W, Tan, and M.E Welge, "Knowledge Management and Data Mining for Marketing", Decision Support Systems, vol. 31. No. 1, 2001, pp. 127-137.

[2] E.W.T. Ngai, L. Xiu, and D. C. K. Chau, "Application of Data Mining Techniques in Customer Relationship Management: A Literature Review on Classification", Expert Systems with Applications, vol.36, no. 2, Part 2, 2009, pp.2592-2602.

[3] S. Mitra, S.K Pal, and P.Mitra, "Data Mining in Soft Computing Framework: A Survey", Neural Networks, IEEE Transactions, vol. 13, no, 1, 2002, pp. 3-14.

[4] S. Harinath, and S.R. Quinn, Analysis services 2005 with MDX, Wiley Publishing, Inc, Indianapolis, Indiana, 2006.

[5] M.J.A. Berry, and G.S. Linoff, Data Mining Techniques: For marketing, Sales and Customer Relationship Management, Wiley Publishing, Inc; Indianapolis, 2004.

[6] J. Han, M. Kamber, Data Mining: Concepts and Techniques, Morgan Kaufmann, San Francisco, 2006.

[7] Y. Kim, "Toward a Successful CRM: Variable Selection, Sampling, and Ensemble", Decision Support Systems, vol.41, no. 2, 2006, pp. 542-553.

[8] R. Potharst, U. Kaymak, and W. Pijls, Neural networks for Target Selection in Direct Marketing', in K. Smith and J. Gupta Neural Networks in Business: Techniques and Applications, Idea Group Publishing, London, 2002, pp. 89-110.

[9] C. Rygielski, J. C. Wang, and D.C. Yen, Data Mining Techniques for Customer Relationship Management, Technology in Society, 24, 2002, pp. 483-502.

[10] C. X. Ling, and C. Li, Data Mining for Direct Marketing: Problems and Solutions, Proceedings of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-98), New York, NY, ACM, 1998, pp. 73-79.

[11] S. J. Lee, and K. Siau, A Review of Data Mining Techniques', Industrial Management and Data System, 2001, 101, 1, 41-46.

[12] J. Deichmann, A. Eshghi, D. Haughton, S. Sayek, and Teebagy, N. "Application of Multiple Adaptive Splines (MARS) in Direct Response Modelling", Journal of Interactive Marketing, 2002,16, 4, 15-27.

[13] U. M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy, Advances in Knowledge Discovery and Data Mining, MII Press, Mento Park, 1996.

[14] P. Van der Putten, "Data Mining in Direct Marketing Databases", Complexity and Management: A Collection of Essay, World Scientific. 1999.

Lilian Sing'oei received her Bachelor's degree (Information Technology) from Jomo Kenyatta University of Agriculture and Technology, Kenya in 2007 and is currently pursuing her master's degree (Computer Application Technology) in Central South University, China. She has previously worked with James Finlay (K) Ltd, Kenya (Tea extract), Kenya Power & Lighting CO. Ltd, Kenya and Universal Work, Health & Safety consultancy Ltd in Kenya at the IT Department. She is certified SPSS Statistical analyst. She is currently pursuing Cisco Certified Network Associate (CCNA). She has two papers that are currently under review. Her research interests include

Database Systems, Business Intelligence, Data Mining, Business Information systems, and Internet Technologies.

Jiayang Wang received his Bachelor's, Master's and PHD degrees from Central South University. He is a professor in Central South University of school of Information Science and Engineering. His research interests are data mining and rough sets. He has published many papers in different refereed journals and conferences.