

DIRECT MARKETING WITH THE APPLICATION OF DATA MINING

M Suman

Assoc.professor
KLUNIVERSITY,GreenFields,v
ijayawada,A.P

T Anuradha

Assoc.professor
KLUNIVERSITY,GreenFields,V
ijayawada,A.P

K Manasa Veena

KLUNIVERSITY

ABSTRACT

For any business to be successful it must find a perfect way to approach its customers. Marketing plays a huge role in this. Mass marketing and direct marketing are the two types of it. Mass marketing targets everybody in the society and thus it has less impact on valued customers where direct marketing concentrates mainly on these valued and un loyal customers and promotes only to them which in turn makes profits. For this separation of customers based on their loyalty data mining algorithms and tools are used.

In this paper we discussed the approach of implementation of data mining for direct marketing. We mainly concentrated and studied on why we apply data mining for direct marketing, how we apply and problems one faces while applying data mining concept for direct marketing and the solutions for them in direct marketing.

General Terms

Classification technique, Decision tree, Algorithm.

Keywords

Direct marketing, Data mining,Decision tree.

1. INTRODUCTION

In marketing a product ,there are several forms of sub disciplines, and one of them is direct marketing which involves messages sent directly to consumers usually through email, telemarketing and direct mail.As the traditional forms of advertising (radio, newspapers, television,etc.) may not be the best use of their promotional budgets, many companies or service providers with a specific market use this method of marketing.

1.1 Where we use direct marketing

For example, a company which sells a hair loss prevention product or life insurance policy would have to find a radio station whose format appealed to older male listeners who might be experiencing this problem. There would be no guarantee that this group would be listening

to that particular station at the exact time the company's ads were broadcast. Money spent on a radio spot (or television commercial or newspaper ad) may or may not reach the type of consumer who would be interested in a hair restoring product.

This is where direct marketing becomes very appealing. Instead of investing in a scattershot means of advertising, companies with a specific type of potential customer can send out literature directly to a list of pre-screened individuals. Direct marketing firms may also keep addresses of those who match a certain age group or income level or special interest. Manufacturers of a new dog shampoo might benefit from having the phone numbers and mailing addresses of pet store owners or dog show participants. Direct marketing works best when the recipients accept the fact that their personal information might be used for this purpose. Some customers prefer to receive targeted catalogues which offer more variety than a general mailing.

2. DATA MINING

Data mining is the process of extracting hidden patterns from data. It is commonly used in a wide range of profiling practices, such as marketing, surveillance, fraud detection and scientific discovery.

Generally, data mining (sometimes called data or knowledge discovery) is the process of analyzing data from different perspectives and summarizing it into useful information - information that can be used to increase revenue, cuts costs, or both. Data mining software is one of a number of analytical tools for analyzing data. It allows users to analyze data from many different dimensions or angles, categorize it, and summarize the relationships identified. Technically, data mining is the process of finding correlations or patterns among dozens of fields in large relational databases.

2.1. Data mining in direct marketing

As we already discussed, in Direct marketing, concentrates on a particular group of customers(not loyal and beneficial).So, the data mining technique called the

Supervised Classification is used to classify the customers for marketing.

2.2 The Process

DMT will extract customer data, append it with extensive demographic, financial and lifestyle information, then identify hidden, profitable market segments that are highly responsive to promotions.

2.3 Decision tree

A Decision tree is a popular classification technique that results in flowchart like structure where each node denotes test on a attribute value and each branch represents an outcome of test. The leaves represent classes. Using Training data Decision tree generate 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 computing for calculating probabilities.

Example-

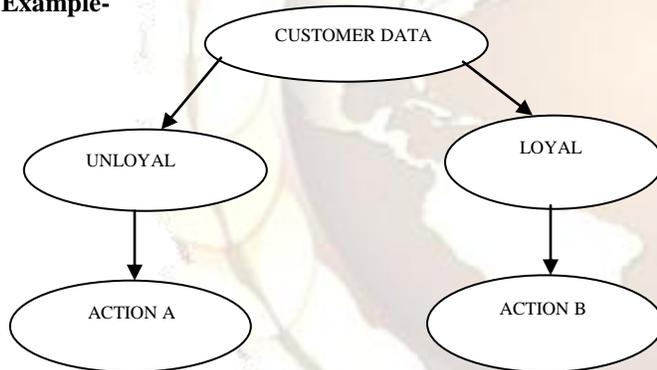


Fig 1:A decision tree based on customer's loyalty

2.4 Building A Decision Tree In Direct Marketing

Decision-tree learning algorithms, such as ID3 or C4.5 are among the most powerful and popular predictive methods for classification. So here in direct marketing we classify the customers on basis of their attributes like sex, age, location, purchase history, feedback details etc.

2.5 Algorithm

C4.5 Builds decision trees from set of training data using the concept of Information entropy.

The training data is a set $S = s_1, s_2, \dots$ of already classified samples. Each sample $s_i = x_1, x_2, \dots$ is a vector where x_1, x_2, \dots represent attributes or features of the sample. The training data is augmented with a vector $C = c_1, c_2, \dots$

where c_1, c_2, \dots 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 subsets. 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.

3. PROBLEMS IN CLASSIFICATION

According to Charles X.Ling and Chenghui Li,the classification of data base involves the following situations:

In the first situation, some (say X%) of the customers in the database have already bought the product, through previous mass marketing or passive promotion. X is usually rather small, typically around 1. Data mining can be used to discover patterns of buyers, in order to single out likely buyers from the current non-buyers, (100-X%) of all customers. More specifically, data mining for direct marketing in the first situation can be discovered in:

1. Get the database of all customers, among which X% are buyers.
2. Data mining on the data set based on Geo-demographic information, transforming address and area codes, deal with missing values, etc.
3. Applying algorithm to prepare objects, classes .
4. Evaluate the patterns formed by applying dmt on testing set.
5. Use the patterns found to predict likely buyers among the current non-buyers
6. Promote to the likely buyers(called rollout).

In the second situation, a brand new product is to be promoted to the customers in the data base, so none of them are buyers. In this case, pilot study is conducted, in which a small portion(say 5%) of the customers is chosen randomly as the target of promotion. Again, X%

of the customers in the pilot group may respond to the promotion. Then data mining is performed in the pilot group to find the likely buyers in the whole database.

Specific problems encountered while data mining on data sets for direct marketing are

1. Imbalance class distribution: Because only a small amount of buyers are likely means positive but most of the algorithms can work on this type of sets. they assume that 100% are unlikely. Many data mining and machine learning researchers have recognized and studied this problem in recent years(Farwett & Provost,19s96;Kubat, Holte, & Matwin; Lewis & Catleltl; Pazzani, Merz, Murphy, Ali, Hume & Brunk).

2. Predictive accuracy cannot be used as a suitable evaluation criterion for the data mining process. Classifying can be difficult. Means considering likely buyers as non-buyers and non-buyers as buyers should be avoided.

4.SOLUTIONS

Ranking of non-buyers makes it possible to choose any number of likely buyers for the promotion.It also provides a fine distinction among chosen customers to apply different means of promotion.

Lift analysis has been widely used in database marketing previously(Hughes,1996).A lift reflects the redistribution of responders in the testing set after the testing examples are ranked.

5. CONCLUSION

Direct marketing is widely used in the fields of marketing like telemarketing,direct mail marketing,email marketing etc.,data mining is applied on this marketing strategy to avoid human flaws in classifying the customers based on their loyalty.We discussed the problems one faces in applying the datamining for direct marketing and discussed their solutions.

6. ACKNOWLEDGEMENTS

This work was supported by Mrs. T Anuradha Assoc.professor and Mr. M Suman Assoc. Professor ,Department of Electronics and Computer Science Engineering, KLUUniversity.

7. REFERENCES

[1] J.R. Quinlan, Morgan Kaufmann, C4.5 Programs for Machine Learning. 1993.

[2] A.Berson, K. Thearling, and S.J. Smith, Building Data Mining Applications for CRM. McGraw-Hill. 1999.

[3] G.K.Gupta, Introduction to data mining with case study ,Prentice Hall of India.2006.

[4] Mehmed Kantardzic, (2003), Data Mining: Concepts, Models,Methods, and Algorithms, John Wiley & Sons.

[5] Farwett & Provost,19s96;Kubat, Holte, & Matwin;Lewis & Catleltl;Pazzani,Merz,Murphy,Ali, Hume & Brunk.

[6]data-mine.com/white_papers/direct_marketing.

[7] The Application of Data Mining For Direct Marketing Purushottam R Patil, Pravin Revankar, Prashant Joshi
Second International Conference on Emerging Trends in Engineering and Technology, ICETET-09.