Customer Relationship Management System Using Decision Tree Classification

Mya Pwint Cho, Ei Ei Han
Computer University, (Pyay)
cholay704@gmail.com,cupyay2009@gmail.com

Abstract

Customer Relationship Management (CRM) has been the important part of enterprise decision and management and Data mining technology provides a good support for the implementation of CRM. This system indents to classify the level of customers according to their prospective level. This prospective level is identified as High, Medium, and Low. The prospective level of customers is determined by Decision Tree Classification of Data Mining. The ability of Data Mining is to extract rules and then deduce useful knowledge automatically from a large number of data collections. Decision trees can easily be converted to classification IF-THEN rules by using decision tree induction. Decision Tree predict more than classifies expert article correctly. This system is implemented for Heavy Machinery Spare Part sales marketing team.

1. Introduction

This paper mainly discusses Customer Relationship Management system by the decision tree classification methods. Customer Relationship Management (CRM) means managing all customers' interactions. This requires using information about the customers to more effectively interact with the customers. In CRM, data mining is frequently used to assign a score to a particular customer or prospect indicating the likelihood that the individual in the wanted way [5].

Data Mining is the process of using computing power to apply methodologies. Data mining can help selecting the targeted customers. This system is implemented for Heavy machinery spare part sales marketing company to analyze future plan but need to get required information. Rather than randomly contacting a customer through a call center or sending mail, a company can concentrate its efforts on the customer who are predicted to have a high likelihood of responding to an offer [4]. The system can analyze to classify the prospective level of customer. If the company may wish to contact the customers about an offer, the company can contact to the targeted customer by using the prospective level of customer. The system can be help effectively in marketing efforts. The most widely used and successful method of classification is C4.5 decision trees [6].

2. Related Work

Data mining can also be helpful to human-resources departments in identifying the characteristics of their most successful employees. Information obtained, such as universities attended by highly successful employees can help HR focus recruiting efforts accordingly. Additionally, Strategic Enterprise Management applications help a company translate corporate-level goals, such as profit and margin share targets, into operational decisions, such as production plans and workforce levels. Data Mining is commonly used in a wide range of profiling practices, such as marketing, fraud detection and scientific discovery [4].

3. Classification

Classification is a data mining technique to predict categorical class labels. It classifies data (constructs a model) based on the training set and the values (class labels) in a classifying attribute and uses it in classifying new data.

Classification is a two step process. First step is Model Construction. In this step, each sample is assumed to belong to a predefined class, as determined by the class label attribute. The set of samples used for model construction: training set. The model is represented as classification rules, decision trees, or mathematical formulae.

Second step is Model Usage. In Second step, the accuracy of the model is estimated. The known label of test sample is compared with the classified result from the model. Accuracy rate is the percentage of test samples that are correctly classified by the model. Test set is independent of training set, otherwise over-fitting will occur [4][5]. If the accuracy of the model is considered acceptable, the model can be used to classify future data samples for which the class label is not known.

4. Algorithm Description

4.1. C4.5 Decision Tree analysis

Decision trees are one of the most popular methods used for inductive inference. They are robust for noisy data and capable of learning disjunctive expressions. The basic algorithm for decision tree induction is a greedy algorithm that constructs decision trees in a top-down recursive divide-and-conquer manner.

4.2. Decision Tree Induction

Let **S** be a set consisting of s data samples. Suppose the class label attribute has **m** distinct values defining **m** distinct classes, $\mathbf{C_i}$. The expected information needed to classify a given sample is given by

$$I(s_1, s_2, \dots, s_m) = -\sum_{i=1}^{m} p_i log_2(p_i)$$

Where $\mathbf{p_i}$ is probability that an arbitrary sample belongs to class $\mathbf{C_i}$ and is estimated by **si/s**. Note that a log function to the base 2 is used since the information is encoded in bits.

Let attribute **A** have **v** distinct values, $\{\mathbf{a_1},\mathbf{a_2},\dots,\mathbf{a_v}\}$. Attribute A can be used to partition **S** into **v** subsets, $\{\mathbf{S_1},\mathbf{S_2},\dots,\mathbf{S_v}\}$, where $\mathbf{S_j}$ contains those samples in S that have value $\mathbf{a_j}$ of **A**. If **A** were selected as the test attribute (i.e., the best attribute for splitting), then these subsets would correspond to the branches grown from the node containing the set **S**. Let $\mathbf{s_{ij}}$ be the number of samples of class $\mathbf{C_i}$ in a subset $\mathbf{S_j}$. The entropy, or expected information based on the partitioning into subsets by **A**, is given by

$$E(A) = \sum_{j=1}^{v} \frac{s_{1j} + \cdots s_{mj}}{s} I(s_{1j}, \dots, s_{mj})$$

The term acts as the weight of the $\mathbf{j^{th}}$ subset and is the number of samples in the subset (i.e., having value $\mathbf{a_j}$ of **A**) divided by the total number of samples in **S** .The smaller the entropy value, the greater the purity of the subset partitions. Note that for a given subset $\mathbf{S_j}$,

$$I(s_{1j}, \dots., s_{mj}) = -\sum_{i=1}^{m} p_{ij} log_2(p_{ij})$$

Where $p_{ij} = \frac{s_{ij}}{|s_j|}$ and is the probability that a sample in $\mathbf{S_j}$ belongs to class $\mathbf{C_i}$. The encoding information that would be gained by branching on **A** is

$$Gain(A) = I(s_1, s_2, \dots, s_m) - E(A)$$

In other words, **Gain (A)** is the expected reduction in entropy caused by knowing the value of attribute **A** [8] [3].

4.3. Decision-tree classification rules

As the name indicates, the decision-tree approach presents a set of decisions in a tree-like structure. Each node specifies a test of some attributes of an instance. Each branch corresponds to one of the possible values of this attribute. Each leaf node represents class label or class label distribution. The basic top-down decision tree generation approach is greedy algorithm. At the beginning, all the training examples are at the root. Then the best attribute for each tree node is selected based on the Relevant Attribute Analysis described the preceding section. During a movement from the root to a leaf node, a node is split into several branches. At each branch, run the algorithm and generate more nodes. The basic steps of the greedy algorithm are summarized below.

(1) The attribute that has the highest information gain is selected as the test attribute of the node.

(2) If all samples are of the same class, then the node becomes a leaf and is labeled with that class; otherwise, the heuristic algorithm selects the attribute with the highest information gain among the others that are not labeled yet. This attribute becomes the "test" or "decision" attribute at the node.

(3) A branch is created for each known value of the test attribute, and the samples are partitioned accordingly.

(4) This recursive partitioning process stops only if one of the following conditions is true: (i) all samples for a given node belong to the same class; (ii) there are no remaining attributes on which the samples may be further partitioned; and (iii) there are no samples for the branch.

The knowledge represented in decision-tree induction can be extracted and represented in a form of IF-THEN rules. One rule is created for each path from the root to a leaf node. Each attribute-value pair along a given path forms a conjunction in the rule antecedent ("IF" part). The leaf node holds the class prediction, forming the rule consequent ("THEN" part) [7].

4.4. Classifier Accuracy

Estimating classifier accuracy is important in that it allows one to evaluate how accurately a given classifier will label future data, that is, data on which the classifier has not been trained. There have techniques for estimating classifier accuracy, such as the holdout and k-fold cross-validation methods.

5. Explanation of the system

Customer Relationship Management System is used for Heavy Machinery Spare Part Sales Marketing Team. The system classifies the customer's prospect level. The prospect level is identified into High, Medium, and Low. This prospect level can be used in decision making for promotion, sales, advertising the new goods, etc.Eg; if the customer has high prospect level, the company will make decision to contact him for marketing. So it can help effectively in marketing efforts.

The prospect level is classified by using decision tree classification. We use the data set of customers. The data set contains two parts: attributes and associated class. . Classification technique divides this data set into training data set and testing data.

5.1. Decision Tree Induction

We use the training data set to analyze decision tree induction. The decision tree induction produces the rules that are needed for the statistical information. The rules support to reduce the lot of condition. We can construct the decision tree as the top down processing.

In the training data set, there are 4 attributes. They are Business Type, Previous Sales, Credit Rating, and Machine Population Status. Each attribute has associated values. Business Type has 2 values: mining and contractor. Previous Sales has yes and no. Credit Rating has excellent, fair, and bad. Machine Population Status has no, local, and large. Class in the data set is defined as Prospect Level. Class has 3 values: High, Medium, and Low. The following Table 1 is the example of the training data set.

Table 1.Prospect level of Customers Table

Business Type	Pre_ Sales	Credit-rating	Mach _Pop_ Status	Prospect Level
Mining	no	excellent	no	high
Mining	yes	excellent	local	high
Contractor	yes	excellent	local	high
Contractor	no	excellent	no	Low
Mining	yes	excellent	local	high
Mining	no	fair	local	medium
Contractor	yes	fair	large	medium
Contractor	yes	bad	large	Low
Mining	yes	bad	local	Low
Contractor	yes	bad	large	Low

The abbreviations of Prospective level of customer table are specified as follows:
Pre_Sales =Previous Sales
Mach_Pop_Status=Machine Population Status

Assumptions:

- Class H : Prospective level=High
- Class M : Prospective level=Medium
- Class L : Prospective level=Low

Information needed to classify a given sample:
I (H, M, L) =I (4, 2, 4) =1.522

Table.2. Entropy for Decision Tree Induction

Credit Rating	H_i	M_i	L_i	$I(H_i, M_i, L_i)$
Excellent	4	0	1	0.722
Fair	0	2	0	0
Bad	0	0	3	0

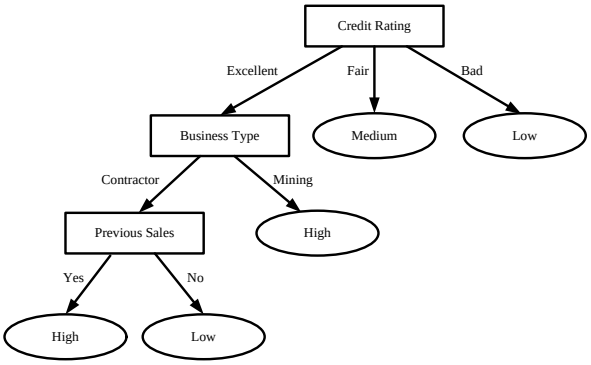

Figure.1.Decision Tree for prospect level

Now E (Credit Rating) =0.361
Hence Gain (Credit Rating) =1.161
Similarly Gain (Business Type) =0.239
Gain (Previous Sales) =0.033
Gain (Machine Population Status) =0.361

As seen from the decision tree in Figure 1, there are a total of five paths in the tree, indicating that five Classification Rules can be extracted, which are stated below. These rules will be used for prediction in next step.

If Credit Rating = fair
then Prospective level = medium.
If Credit Rating = Bad
then Prospective level = low.
If Credit Rating = Excellent
and Business type = Mining
then Prospective level = High.
If Credit Rating = Excellent
and Business type =Contractor
and Previous Sales = No
then Prospective level = low.
If Credit Rating = Excellent
and Business type = Contractor
and Previous Sales = Yes
then Prospective level = High.

5.2. Classifier Accuracy

In this system, classifier accuracy is compute by hold out method. The total data set record is 500 records in the system. The hold out method reserves a certain amount of data for testing and uses the remainder for training-so they are disjoint. One third of data is used for testing, and the rest for training. So the testing record is 167 records and the remainder record is the training data set. A training sample is used for generating the classification model. A test sample is used for the resulting classification model in terms of accuracy.

In addition, the accuracy for measure the result is computed as follows. Accuracy is defined as:

$$\text{Accuracy} = \frac{\text{numbers of correct count} \times 100}{\text{numbers of total records}}$$

Example:
Testing Rules (testing record #1) =record#1.class-
Testing Rules (testing record#2) not=record#2.class-
Testing Rules (testing record #3) =record#3.class-
Testing Rules (testing record#4) =record#4.class-
Testing Rules (testing recodr#5) not=record#5 class-
Error Records:
2errors: #2 and #5
Correct Record = Total records – Error Records
=5-2
=3
Accuracy = 3/5×100=60%

6. System Flow Diagram

This paper will provide Customer Relationship Management system. There are two types of users: Admin and Guest. Admin can manage the training data set, constructs the model, generate decision tree, generate rules, test the accuracy of these rules and then classify the prospective level of customer. Guest can only enquiry the prospective level of customers.

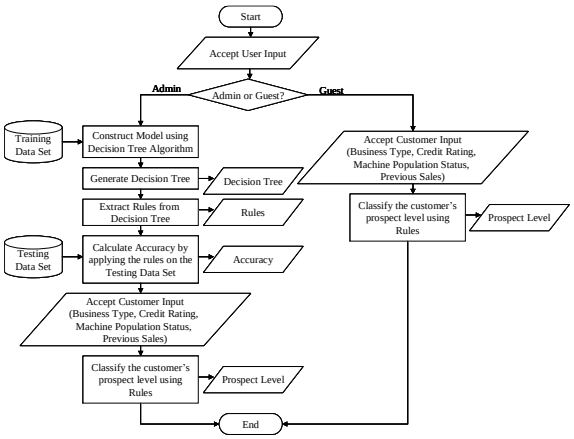

Figure.2.System Flow Diagram

7. Conclusion

In this paper, Customer Relationship Management System that displays the Prospective Level of customer is described by using Decision tree classification.CRM helps companies improve the profitability of their interactions with customers while at the same time making the interactions appear friendlier through individualization. The database of the proposed system is used MS.Access and VB.Net is used. The Prospective level of customer can help the company for decision making to contact customers. This system will effectively help in marketing efforts.

8. References

[1] A.L. Blum,, and P. Langley, (1997). Selection of Relevant Features and Examples in Machine learning. Artificial Intelligence, 97, pp. 245-271.

[2] C. Apte and S. Weiss. Data mining with decision trees and decision rules. Future Generation Computer Systems, 13, 1997.

[3] C. Cardie, Using Decision Trees to Improve Case-based Learning. ICML 1993, pp. 25-32.
.
[4] Ellen Monk, Bret Wagner (2006).”Concepts in Enterprise Resource Planning, Second Edition.”Thomson Course Technology, Boston, MA.

[5] Herb Edelstein, President, Two Crows Corporation.”Building Profitable Customer Relationships with Data Mining”

[6] J.R. Quinlan, (1993). C4.5: Programs for Machine Learning. San Mateo, CA: Morgan Kaufmann

[7] Kejun Fu, Dianming Geng, Department of Business and Management, Shandong Instititude of Business and Technology, Yantai, 264005,China. Determining the Targeted Customers Based on Data Mining

[8] L.Breian, J.Friedman, R.Olshen, and C.Stone. Classification and Regression Tree, Wadsworth International Group, 1984.