ISSN 2320 2629
Volume 3, No.1, January – February 2014
International Journal of Information Technology Infrastructure
Available Online at http://warse.org/pdfs/2014/ijiti02312014.pdf

TECHNIQUES USED IN DECISION SUPPORT SYSTEM FOR CRM- A REVIEW

Ms. Simi Sojan[1], Ms.Sify K Raphy[2],Ms.Pinky Thomas[3]
[1]Department of Computer Science and Engineering
Jyothi Engineering College, Cheruthuruthy, Thrissur, India
simikeettikkal@gmail.com

[2]Department of Computer Science and Engineering
Jyothi Engineering College, Cheruthuruthy, Thrissur, India
sifykarippery@gmail.com

[3]Department of Computer Science and Engineering
Jyothi Engineering College, Cheruthuruthy, Thrissur, India
pinkymukalal@gmail.com

ABSTRACT

A decision Support System (DSS) is a computerized application or a information system that supports business decision more easily. Customer Relationship Management is a strategy which a company uses to handle customer interactions. CRM mainly focuses on attracting, retaining, developing and identifying customers. For Successful CRM approach there we have various data mining algorithms. Here we discuss some algorithms which are used to generating data mining process that are used to implement decision support system for customer relationship management. Using this approach provide an opportunity for organizations to improve their current marketing and CRM.

Keywords: Data mining, Decision Support System, Customer Relationship Management, Clustering, Classification, Association Rule Mining.

1. INTRODUCTION

Today, the companies focus on the goal of understanding each customer individually and using this, it becomes easier for them to improve business. Now a days, the availability of large volume of data on customers, can be effectively used by new information technology tools, to understand customer. Data mining (DM) is the field of computer science that helps in extracting useful information from this large amount of data. Thus it plays an important role in rising firm's profit by improving marketing and Customer Relationship Management (CRM) [6]. The application of different data mining technique commonly used in mining process is applied in CRM domain. The application of data mining in the customer relationship management are, customer classification analysis, customer gaining analysis, customer losing and maintaining analysis, customer profit-making ability analysis and forecast, cross selling analysis, customer satisfaction analysis and customer credit analysis.

A data mining application usually starts with an understanding of the application domain by data analysis, where suitable data sources and target data is then identified. With these data, data mining can be performed.

Here we are discussing about the techniques that are required for extracting information which are useful in Customer Relationship Management. In this paper, the increased amount of information requires advance processing technology as data mining. It creates rules to decide how to retain and develop profitable customers and identify their purchasing habits. Then classifies customers based on some attributes and generate rules to set campaigns for products.

Here, we describe some of the data mining models such as Association, Classification, Clustering and Prediction. In this paper we are going to apply data mining and statistical algorithms those are K-means clustering, classification algorithm, decision tree algorithm (ID3).

2. DATA MINING MODELS

The data mining models that are describing here are the following.

2.1. Clustering

Clustering is a process of partitioning a set of data (or objects) into a set of meaningful sub-classes, called clusters. It helps the users to understand the natural

grouping or structure in a data set. Clustering is useful in a number of tasks such as classification, aggregation and segmentation. As clustering is proceeded, interesting object groups may be discovered such as the groups of insurance policy holders with a high average claim cost or the groups of clients in a banking database having a heavy investment in real estate.

By using the clustering algorithm we can divide the data sample into several numbers of groups as per our requirement. Following are the steps for clustering algorithm [2].

I. Start

II. Enter the Value of 'K' where K is number of group.

III. Run k-means algorithm in retail database.

IV. Get the desire number of cluster sample.

V. Generate the plan as per requirement.

VI. End.

The most distinct characteristic of data mining is that, it deals with very large complex data sets. The data sets to be mined often contain millions of object of various types of attributes or variables. This requires the data mining operations and algorithms to be scalable and capable of dealing with different types of attributes. However, most algorithms currently used in data mining do not scale well when applied to very large data sets because they were initially developed for other applications than data mining which involve small data sets. In terms of clustering, we are interested in algorithms which can efficiently cluster large data sets containing both numeric and categorical values because such data sets are frequently encountered in data mining applications. The *k*-means algorithm is well known for its efficiency in clustering large data sets.

Given a set of numeric objects $\boldsymbol{X}$ and an integer number $k(\leq n)$, the *k*-means algorithm searches for a partition of $\boldsymbol{X}$ into *k* clusters that minimizes the within groups sum of squared errors (WGSS). This process is often formulated as the following mathematical program problem *P* (Selim and Ismail, 1984; Bobrowski and Bezdek, 1991)[4]:

Minimize

$$P(W, Q) = \sum_{l=1}^{k} \sum_{i=1}^{n} w_{i,l} d(X_i, Q_l)$$

Subject to

$$\sum_{l=1}^{k} w_{i,l} = 1, \quad 1 \leq i \leq n$$

$$w_{i,l} \in \{0,1\}, \quad 1 \leq i \leq n, 1 \leq l \leq k$$

Where W is an n × k partition matrix, $Q = \{Q_1, Q_{2,\ldots}\}$ g is a set of objects in the same object domain, and d(., .) is the squared Euclidean distance between two objects.

2.2. Classification

Classification is a supervised algorithm. The classification algorithm predicts categorical class labels. It classifies data (construct a model) based on the training set and the values in a classifying attribute and uses it in classifying new data.

Supervised learning has been a great success in real-world applications. It is used in almost every domain, including text and Web domains. Supervised learning is also called classification or inductive learning in machine learning. This type of learning is analogous to human learning from past experiences to gain new knowledge in order to improve our ability to perform real-world tasks. However, since computers do not have experiences, machine learning learns from data, which are collected in the past and represent past experiences in some real-world applications [7].

A data set used in the learning task consists of a set of data records, which are described by a set of attributes $A = \{A1, A2, \ldots, A|A|\}$, where $|A|$ denotes the number of attributes or the size of the set *A*. The data set also has a special target attribute *C*, which is called the class attribute. In our subsequent discussions, we consider *C* separately from attributes in *A* due to its special status, i.e., we assume that *C* is not in *A*. The class attribute *C* has a set of discrete values, i.e., $C = \{c1, c2, \ldots, c|C|\}$, where $|C|$ is the number of classes and $|C| \geq 2$. A class value is also called a class label. A data set for learning is simply a relational table. Each data record describes a piece of past experience. In the machine learning and data mining literature, a data record is also called an example, an instance, a case or a vector. A data set basically consists of a set of examples or instances.

Given a data set *D*, the objective of learning is to produce a classification/ prediction function to relate values of attributes in *A* and classes in *C*. The function can be used to predict the class values/labels of the future data. The function is also called a classification model, a predictive model or simply a classifier. It should be noted that the function/model can be in any form, e.g., a decision tree, a set of rules, a Bayesian model or a hyperplane.

In Classification data mining technique we can apply ID3 algorithm to do the prediction of unknown value of attribute. Following are the steps.

I. Start

II. Select the categories which are to be classified.

III. Run the basic classification algorithm.

IV. Get the result.

V. Derive conclusion.

VI. End.

2.3. Association

Association rules are for classification criteria. An association rule is an expression X⇒Y, where X and Y are disjoint itemsets. An association rule must not be considered not only as an implication, but rather as a coexistence of the two itemsets and support is given by the support of the X∪Y itemset. The confidence of an association rule is the conditional probability that a transaction contains Y, given that it contains X, and computed using the formula c(X ⇒Y) = б(X∪Y) /б(X). Minimum confidence of a rule is a user defined value. An association rule is strong if it has a support greater than minimum support value and confidence greater than the minimum confidence value[6].

The process of association rule mining approach usually finds out a huge number of rules. These rules are then pruned down on the basis of their coverage value, which is defined as the number of instances from the whole set, where the rule predicts correctly, and their accuracy. Nowadays, coverage is often called as support and accuracy is called as confidence.

The LHS and RHS of the association rules represents combinations of attribute- value pairs that have predefines minimum coverage. These are called itemsets. An attribute value pair is called an item. It basically comes from the process of Market Basket analysis, where the store manager analysis the different items purchased by the customer in a single purchase, and tries to find out association rules among them [7].

Association rules of the form {X1, X2….Xn} ⇒ Y, meaning that if we find X1,X2….Xn in the market basket, then we have a good chance of finding Y. We normally would search only for rules that has confidence above the threshold and may also ask that the confidence be significantly higher, than it would be if items were placed at random into baskets [6].

Association algorithm is helpful to find frequent item set from database. A frequent itemsets, are the sets of items that have minimum support. The subset of a frequent itemset must also be a frequent itemset. i.e., if {*AB*} is a frequent itemset, then both {*A*} and {*B*} should be a frequent itemset. Iteratively find frequent itemsets with cardinality from 1 to *k (k*-itemset*)*. Use these frequent itemsets to generate association rules.

Following is the steps for Association mining[2].

I. Start.

II. Select the categories.

III. Run apriori algorithm in selected categories.

IV. Get the frequent item set.

V. Generate plan as per business need.

VI. End

2.4. Prediction (Decision Tree)

A decision tree is a tree in which each branch node represents a choice between a number of alternatives, and each leaf node represents a decision [3].

Decision tree are commonly used for gaining information for the purpose of decision -making. Decision tree starts with a root node on which it is for users to take actions. From this node, users split each node recursively according to decision tree learning algorithm. The final result is a decision tree in which each branch represents a possible scenario of decision and its outcome.

Decision trees classify instances by traverse from root node to leaf node. We start from root node of decision tree, testing the attribute specified by this node, then moving down the tree branch according to the attribute value in the given set. This process is the repeated at the sub-tree level.

The following are the steps for prediction [2] .

I. Start

II. Give the name of product which is to be predicted.

III. Apply ID3 algorithm.

IV. Get the result either yes or no

V. End

ID3 is a simple decision tree learning algorithm developed by Ross Quinlan (1983). The basic idea of ID3 algorithm is to construct the decision tree by employing a top-down, greedy search through the given sets to test each attribute at every tree node. In order to select the attribute that is most useful for classifying a given sets, we introduce a metric---information gain. To find an optimal way to classify a learning set, what we need to do is to minimize the questions asked (i.e. minimizing the depth of the tree). Thus, we need some function which can measure which questions provide the most balanced splitting. The information gain metric is such a function [3].

3. CONCLUSION

The hike in the participation of today's business world has caused the data input to increase at a tremendous rate thus making data mining a very important part of industries. An interest of various companies has been attracted more and more due to the implementation of continuous Data mining processes.

Here we have proposed a novel data mining process model for marketing and CRM requires. The model endeavors to separate autonomous mining sub-processes from a full DM process, and form session mining and merge mining. The formal framework on data mining based on a session model. Its goal is to develop an autonomous, continuous mining process model for marketing management. The approach represents an opportunity for organizations to improve their current marketing and CRM. Increased efficiency will be realized by the proper alignment of an organization's skills with the task at hand.

REFERENCES

[1] Md. Rashid Farooqi ,Khalid Raza***,"A Comprehensive Study of CRM through Data Mining Techniques"***, Proceedings of the National Conference; NCCIST-2011 September 09, 2011.

[2] Sunil Yadav, Aadithya Desai, Vandana Yadev, ***"Knowledge Management in CRM using Data mining Techniques"***,International Journal of Scientific & Engineering Research.

[3] Wei Peng, Juhua Chen, Haiping Zhou***" An Implementation of ID3 --- Decision Tree Learning Algorithm",*** Project of Comp 9417: Machine Learning University of New South Wales, School of Computer Science & Engineering, Sydney, NSW 2032, Australia.

[4] Zhexue Huang***," Extensions to the k-Means Algorithm for Clustering Large Data Sets with Categorical Values",*** Data Mining and Knowledge Discovery 2, 283–304 (1998)1998 Kluwer Academic Publishers. Manufactured in The Netherlands.

[5] Rakesh Agarwal,Tomasz Imielinski, Arun Swami,***"Mining Association Rules between Sets Of Items in Large Database"***,IBM Almaden Research Center 650 Harry Road,San Jose,CA 95120.

[6**] *"Data Mining Process Model For Marketing And Crm Ding Pan"***, School of Management, Jinan University, July 2010

[7] Bing Liu,"**Web DataMining"** Exploring Hyperlinks, Department of Computer Science University of Illinois at Chicago 851 S. Morgan Street Chicago, IL 60607-7053 USA liub@cs.uic.edu