Special Issue Published in International Journal of Trend in Research and Development (IJTRD), ISSN: 2394-9333, www.ijtrd.com

A Novel Approach to Data Discretization Using Clustering Techniques in Mining of High Dimensional Data

[1]S.Muthukumaran, [2]A. Victoria Anand Mary, [3]M.Valli
[1]Research Scholar, Research and Development Centre, Bharathiar University, Coimbatore, India
[2]Assistant Professor, St.Joseph's College of Arts & Science (Autonomous), Cuddalore, TamilNadu, India
[3] Research Scholar, The KarpagamUniversity, Coimbatore, TamilNadu, India

***Abstract*--** Data preprocessing is a vital step in data mining. Preprocessing resolves various types of data discrepancies encountered in large databases in order to produce quality data for the mining task. Data preprocessing includes four fundamental steps namely data cleaning, integration, reduction and transformation. There are various techniques involved in each step of Data preprocessing. In order to develop quality data, a data miner must decide the most appropriate techniques in every step of preprocessing. In this paper we focus on data reduction, particularly data discretization as one of the most important preprocessing step. Data reduction involves reducing the data distribution by reducing the range of continuous data into a range of values or categories. Data discretization plays a major role in reducing the attribute intervals of data values. Finding an appropriate number of discrete values will improve the performance of data mining modeling, particularly in terms of classification accuracy. This research paper proposes four levels of data discretization taxonomy as follows namely (i) hierarchical and non-hierarchical; (ii) splitting, merging, and combination; (iii) supervised and unsupervised combinations; (iv) binning, entropy and chi-square merge techniques.

***Keywords*--** *Data Preprocessing, Data reduction, Data Discretization.*

I. INTRODUCTION TO DATA DISCRETIZATION

It is a data reduction approach that transforms continuous attribute into discrete attributes. It is used to reduce the total data volume of continuous attributes. Data discretization can also be defined as a process used to quantify continuous attributes. Existing classification tasks cannot be applied continuous attributes as long as the continuous attributes are not discretized beforehand. The use of continuous attributes requires large storage and longer rule. A discretization technique is required to change the continuous attribute to discrete attributes. The use of discrete attributes can increase the accuracy of prediction. [6]The discretization process involves the partition of continuous attributes values into several intervals. Label intervals are used instead of the continuous value of the actual data. Discretization of data increases the accuracy of learning and increases the speed and produce results that are more compact. Discrete attributes are usually more easily interpreted and understandable

A. Factors Involved In Choosing A Discretization Approach

There are mainly four factors involved in choosing a discretization approach. The First factor is the availability of the domain expert from whom several preliminary parameters can be obtained. Second factor is whether the data to be discretized contain the class or target attribute. it is also based on the nature, type and range of distinct values in the data. Third factor is, how the measurement of the data can be applied. Fourth factor is, whether the data is well distributed among the attributes.

B. A Novel Approach to Data Discretization

Many discretization techniques require several parameters to perform the discretization process. These parameters can be predetermined by a data domain expert or they can be automatically determined through training the example data. When the parameter is determined by a domain expert it is called static technique. Dynamic technique is the search for a k value through all possible space for all attributes simultaneously. So that the dependence in attribute discretization is traceable. Local approach discretizes data in a local region of data training space which allows different intervals sets to be performed on an attribute. Global approach discretize data by considering the overall training space and it implements a training process only once.[2] In univariate and unsupervised approach each attribute is considered in isolation and no knowledge of any outcome or decision attribute is employed in this process. In univariate and supervised approach only one condition attribute is considered at a time, but is done so in conjunction with the decision attribute. In Multivariate and supervised approach all condition attributes are considered simultaneously and are done in conjunction with the decision attribute[1].

C. Discretization Methods

a. Common properties of Discretization methods

This section provides a framework for the discussion of the discretizers presented in the next subsection. The issues discussed include several properties involved in the structure of the taxonomy, since they are exclusive to the operation of the discretizer. Other less critical issues such as parametric properties or stopping conditions will be presented although they are not involved in the taxonomy. Finally some criteria will also be pointed out in order to compare discreitzation methods[2].

b. Main characteristics of a Discretizer

The novel approach proposed will be based on these characteristics:

1. Static vs Dynamic:

This characteristic refers to the moment and independence which the discreitzer operates in relation with the learner. [8]A dynamic discreitizer acts when the learner is building the model, thus they can only access partial information embedded in the learner itself, yielding compact and accurate results in conjunction with the associated learner. Otherwise, a static discretizer proceeds prior to the learning algorithm. Almost all known discretizers are static due to the fact that most of the dynamic discretizers are really subparts or stages of Data Mining algorithms when dealing with numerical data.

2. Univariate and Multivariate:

Multivariate techniques also known as 2D discretization, simultaneously consider all attributes to define the initial set of cut points or to deicide the best cut point altogether. They can also discretize one attribute at a time when studying the interactions with other attributes, exploiting high order relationships, By contrast, univariate discretization scheme in each attribute remains unchanged in later stages.

3. Supervised vs Unsupervised:

Unsupervised discretization do not consider the class label whereas the supervised ones do. The manner in which the latter consider the class attribute depends on the interaction between input attributes and class labels, and the heuristic measures used to determine the best cut points (entropy, interdependence etc). [9]Most discretizers proposed in the literature are supervised and theoretically using class information should automatically determine the best number of intervals for each attribute. If a discreitzer us unsupervised, it doesn't mean that it cannot be applied over supervised tasks. However a supervised discreitzer can only be applied over supervised DM problems.

4. Splitting vs Merging:

This refers to the procedure used to create or define new intervals. Splitting methods establish a cut point among all the possible boundary points and divide the domain into two intervals. By contrast merging methods start with a pre-defined partition to mix both adjacent intervals. These properties are highly related to Top-Down and Bottom-up respectively. The idea behind them is very similar, except that top-down or bottom-up discretizers assume that the process is incremental, according to a hierarchical discretization construction. In fact there can be discretizers whose operation is based on splitting or merging more than one interval at a time.

5. Direct vs Incremental:

Direct discreitzers divide the range into k intervals simultaneously, requiring an additional criterion to determine the value of k they do not only include one-step discreitzatio methods, but also discretizers which perform several stages in their operation, selecting more than a single cut point at every step, by contrast incremental methods begin with a simple discreitzation and pass through an improvement process. requiring an additional criterion to know when to stop it. At each step, they find the best candidate boundary to be used as a cut point and afterwards the rest of the decisions are made accordingly. Incremental discretizers are also known as hierarchical discretizers. Both types of discretizers are widespread in the literature, although there is usually a more defined relationship between incremental and supervised ones.

D. Evaluation Measure

This is the metric used by the discretizer to compare two candidate schemes and decide which is more suitable to be used. We consider five main families of evaluation measures:

a. Information

This family includes *entropy* as the most used evaluation measure in discretization (MDLP, ID3, FUSINTER) and other derived information theory measures such as the *Gini index*.

1. ***b. Statistical***

2. Statistical evaluation involves the measurement of dependency/correlationamong attributes (Zeta, ChiMerge, Chi2), probability and bayesian properties (MODL), interdependency, contingency coefficient, etc.

3. ***c. Rough Sets***

4. This group is composed of methods that evaluate the discretization schemes by using rough set measures and properties such as lower and upper approximations, class separability, etc.

5. ***d. Wrapper***

6. This collection comprises methods that rely on the error provided by a classifier that is run for each evaluation. The classifier can be a very simple one, such as a majority class voting classifier (Valley) or general classifiers such as Naive Bayes (NBIterative).

7. ***e. Binning***

8. This category refers to the absence of an evaluation measure. It is the simplest method to discretize an attribute by creating a specified number of bins. Each bin is defined apriori and allocates a specified number of values per attribute. Widely used binning methods are Equal-Width and Equal-Frequency..

f. Parametric vs. Non-Parametric:

This property refers to the automatic determination of the number of intervals for each attribute by the discretizer. A nonparametric discretizer computes the appropriate number of intervals for each attribute considering a trade-off between the loss of information or consistency and obtaining the lowest number of them. A parametric discretizer requires a maximum number of intervals desired to be fixed by the user. Examples of nonparametric discretizers are MDLP and CAIM. Examples of parametric ones are ChiMerge and CADD .

g. Top-Down vs. Bottom Up:

This property is only observed in incremental discretizers. Top-Down methods begin with an empty discretization. Its improvement process is simply to add a new cutpoint to the discretization. On the other hand,[19] Bottom-Up methods begin with a discretization that contains all the possible cutpoints. Its improvement process consists of iteratively merging two intervals, removing a cut point. A classic Top-Down method is MDLP and a well-known Bottom-Up method is ChiMerge.

h. Stopping Condition:

This is related to the mechanism used to stop the discretization process and must be specified in nonparametric approaches. Well known stopping criteria are the Minimum Description Length measure, confidence thresholds, or inconsistency ratios.

E. Criteria to Compare Discretization Methods

When comparing discretization methods, there are a number of criteria that can be used to evaluate the relative strengths and weaknesses of each algorithm. These include the number of intervals, inconsistency, predictive classification rate and time requirements.

a. Number of Intervals

A desirable feature for practical discretization is that discretized attributes have as few values as possible, since a large number of intervals may make the learning slow and ineffective.

b. Inconsistency

[11]A supervision-based measure used to compute the number of unavoidable errors produced in the data set. An unavoidable

error is one associated to two examples with the same values for input attributes and different class labels. In general, data sets with continuous attributes are consistent, but when a discretization scheme is applied over the data, an inconsistent data set may be obtained. The desired inconsistency level that a discretizer should obtain is 0.0.

c. Predictive Classification Rate:

A successful algorithm will often be able to discretize the training set without significantly reducing the prediction capability of learners in test data which are prepared to treat numerical data.

A static discretization process is carried out just once on a training set, so it does not seem to be a very important evaluation method. However, if the discretization phase takes too long it can become impractical for real applications. In dynamic discretization, the operation is repeated many times as the learner requires, so it should be performed efficiently.

II. ATTRIBUTE SUBSET SELECTION

Feature selection is a process that selects a subset of original features. [14]Feature selection is one of the important and frequently used techniques in data preprocessing for data mining. In real-world situations, relevant features are often unknown a priori. Hence feature selection is a must to identify and remove are irrelevant/redundant features. It can be applied in both unsupervised and supervised learning. [11]The goal of feature selection for unsupervised learning is to find the smallest feature subset that best uncovers clusters form data according to the preferred criterion. Feature selection in unsupervised learning is much harder problem, due to the absence of class labels. Feature election for clustering is the task of selecting important features for the underlying clusters[3].

The subset selection is reduces the dimensionality of the data and enables learning the data faster and more effectively. Generally, attributes are classified attributes are classified as:

Relevant: These are attributes have an influence on the output and their role cannot be assumed by the rest.

Irrelevant: Irrelevant attributes are defined as those not having influence on the output, and whose values are generated at random for each example.

Redundant: A redundancy exists whenever a attribute can take the role of another.

A. Why We Select / Extract Features:

1. To improve accuracy
2. Reduce computation
3. Reduce space
4. Reduce cost of future measurements
5. Improved data/model understanding

B. Forward Selection:

1. start with no features
2. try each feature not used so far in the classifier
3. keep the one that improves training accuracy most
4. repeat this greedy search until all features are used
5. you now have a ranking of the M features and M classifier
6. test each of the M classifier on a validation set return the feature subset corresponding on a validation set.
7. Return the feature subset corresponding to the classifier with lowest validation error.

C. Backward Elimination:

1. Start with ALL features
2. Try discarding each feature currently in the classifier
3. Discard the one that causes LEAST decrease in training accuracy
4. Repeat this until only one feature remains[4]

The figure below shows the process of attribute subset selection in discretization. Here the original feature is taken as set and then we have to generate the candidate the subset. The subset has to be evaluated using a proper discretization function. Until the stopping criterion is reached the process has to be repeated[5].

D. Process of Discretization

1. The continuous attribute is taken as the input and sorted.
2. The discretization process selects a candidate as a cut point using adjacent intervals.
3. It invokes an appropriate measure and it splits or merge based on the measure.
4. This measure continuous until the stopping criterion is reached.
5. The stopping criterion controls the overall discretization process.

E. Proposed Algorithm for Discretization:

i. Let D be a dataset contain all the features of a Data table, sort the data in ascending order and split D into train| validation| test sets Tr.
ii. For each subset, train a classifier using Tr.
iii. Select a appropriate measure to select the candidate cut point(Entropy, Info gain, chi square, karl's spearsman rank co-efficient).
iv. When the stopping criterion is reached split/merge based on the adjacent intervals.
v. Return the feature subset Ω corresponding to the classifier with lowest validation error.
vi. Repeat the steps until all features are used.

The figure shown below explain about the process of discretization

III. THE HIERARCHIAL APPROACH

The hierarchical approach can be divided into three sub approaches, namely the Splitting (top-down),Merging (bottom-up)and Combination. In this section various techniques that fall under this categories are discussed[6].

A. The Splitting Approach

The main concept behind the Splitting approach is the creation of interval cut-off points and adds new items to the list by splitting or dividing intervals through the discretization process.

a. Binning

[17]Binning methods smooth a sorted data value by consulting its neighborhood, that is the value around it. The sorted values are divided into a number of buckets or bins. Binning methods consult the neighborhood values and performs local smoothing.

b. 3-4-5 rule:

It is used to segment numerical data into relatively uniform, natural seeming intervals. If an interval covers 3, 6, 9 then partition the range into 3 intervals. (2-3-2 for 7 intervals). If it covers 2,4 or 8 distinct values then partition it into 4 equal-width intervals. If it covers 1, 5 or 10 distinct values then partition the range into 5 equal-width intervals. Eg sorted data for prize in rupees [4,8,15,21,21,24,25,28,34] partition into bins. By applying 3-4-5 rule.

c. Equal- frequency method:

The bins having the equal frequencies.

Bin 1: 4, 8, 15

Bin 2: 21, 21, 24

Smoothing by bin means: By calculating the mean of the bin and apply the mean value to all attribute.

Bin 1: 9, 9, 9

Bin 2: 22, 22, 22

Bin 3: 29, 29, 29

Smoothing by bin boundaries: The minimum and maximum values are identified as bin boundaries. Each bin is replaced by closest bin value.

B. Merging Approach

These Merges are also known as bottom-up discretization. The process begins with a complete list of continuous attributes as the individual cut off points. The process reduces the number of intervals during the discretization process. [11]The merging discretization may occur either in an unsupervised or supervised approach. The unsupervised approach includes the cluster analysis discretization(CA), k-means clustering discretization (k-means). The supervised merging techniques include the chi merge, chi2, stat disc, info merge, off-line discretization(OLD).

a. Interval Merging By χ^2 Analysis:

This is a bottom-up approach by finding the best neighboring intervals and then merging these to form a larger intervals recursively. This method is supervised because it uses the class information. For accurate discretization, the relative class frequencies should be fairly consistent within an interval. If two adjacent intervals have very similar distribution of classes, then the intervals can be merged. Otherwise they should remain separate. Chi-square test is used for the following purposes like, to test the independence of attributes, population variance, and the goodness of fit[7].

Eg: The following data are samples of 300 car owners in which they are classified with respect to age and number of accidents they meet.

The formula for chi-square is

$$\chi 2 = \sum [(O-E)^2/E] \sim (r-1)(c-1)[\text{degree of freedom}]$$

where O is the observed frequency, E is the expected frequency, r-row, c- column. Expected frequency is calculated from the formula E=RT×CT/GT where RT is the row total and CT is the column total and GT is the grand total.

Where O-Observed Frequency and E-Expected Frequency.

b. Hypothesis:

There are two types of hypothesis (i) null hypothesis (ii) Alternate hypothesis. Null hypothesis is denoted by $H_{0:}$ the age and no of accidents are dependent. Alternate hypothesis is denoted by H_1 : the age and number of accidents are not independent or dependent. Table value: the table value can be obtained from chi- square table for the (r-1)(c-1)degrees of freedom. (ie) (3-1)(3-1)=2×2=4. 4 degree of freedom at 5% level of significance the table value is 9.4888. Inference: since the calculated value (16.50) is greater than the table value (9.4888). we reject the null hypothesis (ie) the age and no of accidents they meet are dependent.

C. The Combination Approach

These discretization techniques use both splitting and merging approaches. Splitting can be used to generate an interval while merging can be performed latter. The discretization technique that use both splitting and merging can also be divided into two approaches ie.supervised and unsupervised approaches. [17]Iterative-Improvement Discretization (IID) and Multivariate Discretization (MVD) are the techniques which use unsupervised approach. Cost Sensitive Discretization (CSD) and Cost Based Discretization (CBD) use supervised approach.

IV. NON-HIERARCHIAL TECHNIQUES

The non-hierarchical techniques are the discretization techniques that do not employ hierarchy. These techniques also include unsupervised and supervised approaches. These techniques use method like binning, fuzzy discretization and rough set theory.

A. Fuzzy-Discretizaton Technique

It was developed to generate linguistic association rules. According to association rules continuous attributes must be discretized into appropriate intervals. Most of the linguistic terms cannot be accurately represented by intervals with splitting points. Each continuous value is thus assigned a suitable grade with linguistic terms from the discretization process. This suitable grade is derived mathematically using a membership function in fuzzy logic. The experimental results show that the linguistic rules obtained with fuzzy discretization perform better than the standard association rules in non-fuzzy discretization[8].

B. Discretization Using Rough Set Theory

The proposed method for discretization comprises of discretized intervals by using RST tools. [15]this method uses three threshold values Max-point and Min-point and Max-length. Max-point and minimum point are used as controls on the number of distinct attribute values while max-length is used to limit the range of a normal interval. The outcomes of the clustering phase are categorized as normal, large or small[9].

6. Normal: An interval I is said to be normal if Min-point ≤ Card(I) ≤ Max-point AND Range(I) ≤ Max-length.

7. Large: An interval I is said to be large if Card(I)> Max-point OR Range(I) > Max-Length OR both. interval I is said to be small if Card(I) < Min-point[10].

To achieve good discretization, the partition of discretized intervals needs to be refined by reorganizing the large of small intervals. This process also optimizes the number of intervals by splitting the large and merging the small intervals. The number of class intervals may be defined by using the formula

Number of classes=1+3.322 ($log_{10}{}^{n}$)

where n is the number of observations. Tally bars are used to predict the attribute value..

a. Histogram Analysis

Histogram analysis is an unsupervised discretization technique because it does not use class information. Histogram partition the values for an attribute, A into disjoint ranges called buckets. By using the formula for class interval

$c=1+3.222\log_{10}{}^{n}$ the attributes are partitioned. Tally bars are used to predict the attribute.

b. Types Of Histograms

Equal-width: In an equal width histogram the width of each bucket range is uniform.

Equal-Frequency: the frequency of each bucket is constant.

v-optimal: It is the one with least variance. The histogram variance is a weighted sum of the original values that each bucket represents, where bucket weight is equal to the number of values in the bucket.

Maxdiff: The difference between each pair of adjacent values is considered. A bucket boundary is established between each pair having β-1 largest differences, where β is the user specified number of buckets.

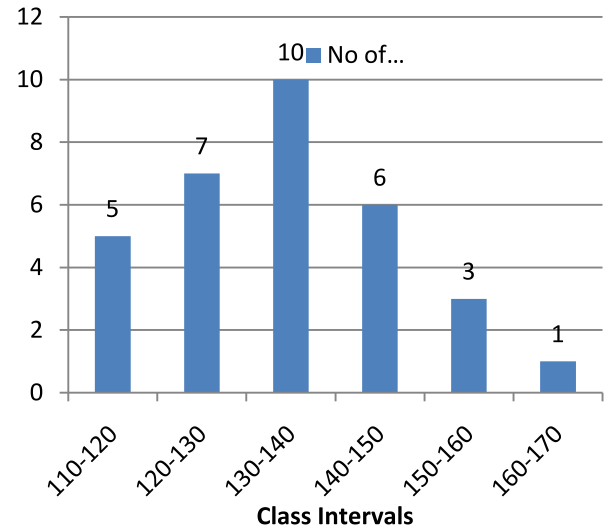

Figure 1: An equal width histogram for the daily wages

C. Entropy Based Discretization

a. Entropy

[13]Entropy uses information gain as its attributes selection measure. The attribute with highest information gain is chosen as the splitting attribute for node N. this attribute minimizes the information needed to classify the examples in the resulting partitions and reflects the least randomness or "impurity" in these partitions[11]. The expected information needed to classify an example in dataset D is given by

$Info(D)=-\sum p_i \log2(p_i)$

Where p_i is the probability that an arbitrary example in dataset D belongs to the class Ci and is estimated by |Ci,D|/|D|. A log function to the base 2 is used. Because the information is encoded in bits. Info(D) is just the average amount of information needed to identify the class label of an examplein dataset D. Partitioning (e.g., where a partition may contain a collection of examples from different classes rather than from a single class) to produce an exact classification of the examples by

$$Info_A(D)=-\sum|D_1|/|D|^{*}\ Info(D)$$

The term $|D_1|/|D|$ acts as the weight of the jth partition $Info_i(D)$ is the expressed information required to classify an example from dataset D based on the partitioning by A. the information gain is defined as the difference between the original information requirement and the new requirement that is.

$$Gain(A)=Info(D)-\ Info_A(D)$$

The attribute A with the highest information gain. Gain(A) is chosen as the splitting attribute at node N.The information gain heuristic adopted in ID3 classifier can be use to find the most informative border to split the value domain of the continuous attribute, when the continuous attribute values in ascending order. The maximum information gain always consider at a cut point or the mid- point between the values taken by the two examples of different classes. Each attribute value of the formula "$A=(A_i+A_{i-1})/2$" where i=I,….n-I is a possible cut point, if A_i and A_{i-1} have been taken by different class values in the dataset. The information gain heuristic check each of the possible cut points and find the best split point. It is a top-down approach and it produces very large number of intervals borders, if the attribute is not very informative.

Let D consist of data tuples defined by set of attributes provides the class information per tuple. Then the entropy discretization method for splitting the attribute A within the set is as follows.

1. Each value of A can be considered asa potential interval boundary or split-point to partition the range of A. A split point for A can partition the tuples in D into subsets satisfying the condition A<split-point and A>split-point.
2. Suppose we have the required information in one class say C1 and have information in another class say C2 and so on.This information is called expected information. When selecting a split-point for attribute A, we want to pick the attribute value that gives the minimum expected information requirement ($\min(Info_A(D))$) by using the formula $Info(D)=-\sum p_i \log2(p_i)$
3. The process of determining a split-point is recursively applied to each partition obtained until some stopping criterion is meet.

D. Background and Related Work

It is found that nearly 45% of the students taken their computer course in SAT scheme which is offered during the summer vacation specially concentrated on 10th and 12th students. 20% of the students take their course at merit offer which is offered to the UG and PG graduates who have finished their academic studies.

The course took by the students are also based on the qualifications of the students. Students who are non-computer professional took the tally,

a. Novel Approach

Since our data are not transactional items but rather relational records. The data obtained from the csc center is maintained in Excel sheet for our purpose we convert the data are converted into Microsoft Acess 2007 and it is the back end for our research. The front end was implemented in .net framework. Connected architecture procedure was used to connect to the database and retrieve the records. The software implementation was done by 2-tier architecture procedure. Before we begin to analyse the data and its attributes we need to perform some preprocessing steps..

V. RESULTS AND DISCUSSION

$Entropy(S)=\sum pi(\log_{2)} pi$ -------(1)

The above equation is used to find the entropy for the data in the table say for example we took scheme and calculate the summation of their weight p times the probability of being in set S. T is the value used to split S into S1 and S2.Entropy(S,T)=|S1|/|S|∑Entropy(S1)+ |S2|/|S|∑Entropy(S2)---------------(2)

In the above data it is observed that the number of students who have took certain diploma course based on gender difference have 95% support and confidence of 3% therefore these two attributes are interconnected with one another. By following the same procedure we took the another attribute for our research for example we took the attribute scholarship awarded.By using the above equation (1) we calculate entropy as Entropy(S) = ((6700/10000 $\log_2$ 6500/10000)-(3300/1000$\log_2$ 3300/10000)) =0.9172.

CONCLUSION

Since data preprocessing is a vital task in any data mining process, we propose a novel approach to data preprocessing using discretization technique for quality data mining. This approach enables the users to select the appropriate technique which suits to their data for a specific domain. As discretization is an important process in the concept of data mining, the selection of appropriate technique will ensure the accuracy of data mining and increase the speed of mining process. Noval approach determines the Finding an appropriate number of discrete values will improve the performance of the data mining modeling, particularly in terms of classification accuracy. [12]This paper proposes four levels of data discretization taxonomy as follows: hierarchical and non-hierarchical; splitting, merging and combination; supervised and unsupervised combinations; and binning, statistic, entropy and other related techniques. The taxonomy is developed based on a detailed review of previous discretization techniques.

References

[1] Sameep Mehta, Srinivasan Parthasarathy, “Toward Unsupervised Correlation Preserving Discretization”, IEEE, 2005.

[2]Salvador Garcia, Julian Luengo, “A Survey of Discretization Techniques: Taxonomy and Emprical Analysis in Supervised Learning”, IEEE 2012.

[3]Xiuqin Ma, Norrozila Binti Sulaiman, “Relation between Significane of Attribute Set and Single Attribute”, World Academy of Science, Engineering and Technology, 2010.

[4]Denis Rutovitz, “Feature Selection-The University of Manchester” May 18, 1996.

[5]Hao Zhang, Duoqian Miao, “A Modified Chi2 Algorithm Based on the Significance of Attribute”, IEEE 2006.

[6]Azuraliza Abu Bakar, “Building a new Taxonomy for data discretization techniques” , IEEE, 2009.
C.R.Gotharie, “Research Methodology”,

[7]Abdelaziz Berrado, Georger C. Runger, “Supervised Multivariate Discreitzation in Mixed Data With Random Forests”, IEEE 2009.

[8]Rough Set-Wikipedia the free encyclopedia.html.

[9]Girish kumar singh and SonajhariaMinz,”Discretization Using Clustering and Rough Set Theory”, IEEE 2007.

[10]Seung-Hyun Kim, Criag Dunhan, SuryoMuljono, “Discovery of Association Rules in National Violent Death Data Using Optimization of Number of Attributes”, IEEE 2009.

[11] Semi-Supervised Learning Literature Survey, X. Zhu, Computer Sciences TR 1530, University of Wisconsin -- Madison.

[12] Breiman, L. and Spector, P. 1992. Submodel selection and evaluation in regression the x-random case. International Statistical Review, 60(3):291–319