ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 3, Issue 2, March 2014

Reducts and Discretization Concepts, tools for Predicting Student's Performance

Jyoti namdeo
Research Scholar, Department of Computer Application, BVUIMED, Pune

Naveenkumar Jayakumar,
Research Scholar, Department of Computer Engineering, BVUCOE, Pune

Abstract— Discretization is a preprocessing task which when conducted leads to very good results in declaring the rules between attributed, classifications of objects and predicting of classes. Various terms like cut-point, arity, binning, splitting method, entropy and rough set has been discussed. Very little work has been reported on discretization of continuous data that too using rough set theory. This research paper has pinned down some of the discretization concepts and tools which are used in relation with the Reducts i.e. Rough Set Theory. This paper also presents the results of the test performed on student's data to discretize the values based on decision attribute and which in turn help in predicting the performance of students.

***Index Terms*— Discretization, Cut-point, Splitting Method, Entropy, Rough Set, student's performance.**

I. INTRODUCTION

Objects in Real world datasets are characterized by attributes which may be nominal, discrete or continuous. The nominal attribute cannot be considered to be associated with an ordered scale for their order is of no consequence whereas discrete values are intervals in a continuous range of values and there are infinitely many values for an attribute which is continuous.

However, various methods in data mining require relatively fever number of attribute values. It is customary to convert input data sets with continuous attributes into input data sets with discrete attributes by partitioning numeric variables into a number of sub-ranges and treat each such sub-range as a category. This process of partitioning continuous variables into categories is usually termed discretization. Discrete values have important roles in data mining and knowledge discovery. They are about intervals of numbers which are more concise to represent and specify, easier to use and comprehend as they are closer to a knowledge-level representation than continuous values. Discretization of real value attributes (features) is an important pre-processing task in data mining for classification problems [11, 14, 26, 27, and 21]. [14] reported that discretization makes learning faster. A hierarchical framework was provided by [21] to categorize the existing numerous discretization methods available in the literatures starting from top down manner as splitting vs. merging, supervised vs. unsupervised, dynamic vs. static, local vs. global and direct vs. incremental.

II. DISCRETIZATION PROCESS

Some terms used in different works followed by an abstract description of a typical discretization process.

A. Feature

"Feature" or "Attribute" or "Variable" refers to an aspect of the data. Usually before collecting data, features are specified or chosen. Features can be discrete, continuous, or nominal.

B. Instance

Instance" or "Tuple" or "Record" or "Data point" refers to a single collection of feature values for all features. A set of instances makes a data set. Usually a data set is in a matrix form where a row corresponds to an instance and a column corresponds to a feature.

C. Cut Point

The term "cut-point" refers to a real value within the range of continuous values that divides the range into two intervals, one interval is less than or equal to the cutpoint and the other interval is greater than the cut-point. For example, a continuous interval [a, b] is partitioned into [a, c] and (c, b], where the value c is a cut-point. Cut-point is also known as split-point.

ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 3, Issue 2, March 2014

D. Arity

The term "arity" in the discretization context means the number of intervals or partitions. Before discretization of a continuous feature, arity can be set to k—the number of partitions in the continuous features. The maximum number of cut-points is k − 1. Discretization process reduces the arity but there is a trade-off between arity and its effect on the accuracy of classification and other tasks. A higher arity can make the understanding of an attribute more difficult while a much lowered value may affect predictive accuracy negatively.

A discretization process broadly consists of four steps [21].

(1) Sorting the continuous values of the feature to be discretized,
(2) Evaluating a cut-point for splitting or adjacent intervals for merging,
(3) Splitting or merging intervals of continuous value,
(4) Finally stopping at some point.

E. Splitting Method

Various discretization measures under splitting are binning [20], entropy [30, 6, 16, 32, 7], dependency [19] and accuracy [8].

F. Binning

In binning measure continuous-valued attribute is discretized by creating a specified number of bins of either equal-width or equal-frequency. Both of these methods are unsupervised whereas 1R [20] is a supervised discretization method using binning.

G. Entropy

Entropy is one of the most commonly used discretization measures. Shannon defines entropy of a sample variable X as [28, 29].

$$H(X) = -\sum_{x} p_x \log p_x$$

Where x represents a value of X and p_x is its estimated probability of occurrence. It is the average amount of information per event where information of an event is defined as:

I (x) = −log p_x

ID3 [30, 31], D2 a successor of ID3 discretization [6], minimum description length principle (MDLP) [16] are some popular algorithms that use entropy measure for discretization.

Dependency Zeta is a measure of strength of association between the class and a feature. [19] defined it as the maximum accuracy achievable when each value of a feature predicts a different class value. .

A Zeta value for a cut-point is:

Z = $\sum_{i=1}^{k} n_{f\,(i).i}$

Where

k = number of prespecified intervals

f (i) = a class index that has the highest count of instances in interval i, and

$n_{f\,(i).i}$ = number of instances in interval i with class index f (i) (modal class index).

Cut-point with the highest Z value is selected if no neighboring pair of partitions predicts the same class. Accuracy measure means the accuracy of a classifier. An example of using accuracy for discretization is Adaptive Quantizer [8]. It considers how well one attribute predicts the class at a time. For each attribute, its continuous range is split into two partitions either by equal-frequency or by equal-width. The splitting is tested by running a classifier to see if the splitting helps improve accuracy [3].

III. MERGING METHODS

This method of discretization is also known as bottom-up method.

A. χ2 measure

χ2 is a statistical measure that conducts a significance test on the relationship between the values of a feature and the class [22]. χ2 statistic determines the similarity of adjacent intervals based on some significance level. It tests the hypothesis that two adjacent intervals of a feature are independent of the class. If they are independent, they should be merged; otherwise they should remain separate.

ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 3, Issue 2, March 2014

B. ChiMerge

It is a supervised, bottom-up discretization procedure [22]. Initially each distinct value of the attribute is considered to be one interval. $\chi 2$ tests are performed for every pair of adjacent intervals. Adjacent intervals with the least $\chi 2$ value are merged together till the chosen stopping criterion satisfies.

C. ConMerge

A method very similar to Chi2 is ConMerge [33]. It also uses the $\chi 2$ statistic and the inconsistency measure. Instead of considering one attribute at a time, ConMerge chooses the lowest $\chi 2$ value among the intervals of all continuous features. [4] Proposes the discretization method Khiops, based on the chi-square statistic. In contrast with related methods ChiMerge and ChiSplit, this method optimizes the Chi-square criterion in a global manner on the whole discretization domain and does not require any stopping criterion. The Khiops method starts the discretization from the elementary single value intervals. It evaluates all merges between adjacent intervals and selects the best one according to the chi-square criterion applied to the whole set of intervals. The stopping rule is based on the confidence level computed with the chi-square statistic. The method automatically stops merging intervals as soon as the confidence level, related to the Chi-square test of independence between the discretized attribute and the class attribute, does not decrease anymore [23].

D. Adaptive Discretization and Evolutionary based methods

In Adaptive Discretization Intervals (ADI) [2] rules are used that contain intervals which are built joining together the low level intervals provided by the discretization algorithm, thus collapsing the search space when it is possible. ADI representation can use several discretization algorithms at the same time allowing the system to choose the correct discretization for each problem and attribute. The authors of [1] generalize the ADI representation approach (proposing ADI2) by also using heuristic non-uniform discretization methods. This representation evolves rules that can use multiple discretizations, letting the evolution choose the correct discretization for each rule and attribute. Moreover, the intervals defined in each discretization can split or merge among them through the evolution process, reducing the search space where it is possible. There are other systems like ADI, perform evolutionary induction of rules based on discretization [18, 12].

The authors of [13] analyzed experimentally discretization algorithms for handling continuous attributes in evolutionary learning. They consider a learning system that induces a set of rules in a fragment of first-order logic, and introduce a method where a given discretization algorithm is used to generate initial inequalities, which describe sub-ranges of attributes' values. Mutation operators exploiting information on the class label of the examples are used during the learning process for refining inequalities.

Many authors proposed multivariate methods that search for cut points simultaneously [24]. [17] proposed an extension to the multivariate case, relying on the multivariate definition of discrete neighborhood by means of a non-oriented graph structure. A framework for supervised bipartitioning has been proposed, which applied recursively leads to a new multivariate discretization algorithm. Non-Disjoint Discretization (NDD) forms overlapping intervals for a numeric attribute always locating a value toward the middle of an interval to obtain more reliable probability estimation [34]. It also adjusts the number and size of discretized intervals to the number of training instances, seeking an appropriate trade-off between bias and variance of probability estimation. [7] proposed a discretization method based on a distance metric between partitions that can be easily implemented in parallel. This method is very effective and efficient in very large datasets [1].

All the above stated methods of discretization are for the data which is certain or crisp. However, real world data possess uncertainty due to data acquisition device error, approximate measurement, sampling fault, transmission latency, data integration error and so on. The concept of uncertainty had been studied by many researchers, and the fuzzy set and rough set [35] emerged as the solution to the problem of uncertainty. It is extremely important to consider data uncertainty in the discretization methods as well. [11] Proposed a global discretization method, based on cluster analysis of transforming any local discretization method into a global one. To measure the consistency for inconsistent data sets, they used a measure called level of consistency based on rough set theory to deal with uncertainty, introduced by [35].A level of consistency, denoted by Lc, is defined as follows

$$L_c = \frac{\sum_{x \in \{d\}^*} |\underline{A}X|}{|U|}$$

ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 3, Issue 2, March 2014
U denote the set of all examples of the data set
P denotes a nonempty subset of the set of all variables, i.e., attributes and a decision.
P* is a partition on U,
Let X is any subset of U.
A is set of all attributes.
Lower approximation of X in P, denoted by $\underline{P}$X and defined as follows
$\cup\{Y \in P^* | Y \subseteq X\}$.
Upper approximation of X in P, denoted by
$\bar{P}$X and defined as
$\cup\{Y \in P^* | Y \cap X \neq \Phi\}$.

A data set is consistent with respect to decision d if and only if A* $\leq$ {d}*, where A is the set of all attributes. In rough set theory, the level of consistency is known as the degree of dependency of {d} from A. For a data set that is consistent with respect to decision d, L_c = 1.
Let A be a continuous attribute, and let the domain of A be the interval [a, b]. A partition π_A on [a, b] is defined as the following set of k subintervals
π_A = {[a0, a1), [a1, a2)... [ak–1, ak]},
Where a0 = a, ai–1 < ai, for i = 1, 2... k, and ak = b.
Thus, discretization is the process that produces a partition π_A on [a, b].

[11] Experimented with three known local methods of discretization as Equal Interval Width Method by [8], Equal Frequency per Interval Method by [33] and Minimal Class Entropy Method by [16]. They reported that Local methods suffer from the inability to predict how many intervals should be induced for a given domain of a continuous attribute A. Being unaware of how many intervals to induce, which method to use, and repeating this step n times (for each continuous attribute) can seriously jeopardize the outcome of the discretization process.

For successful discretization they assumed following guidelines:
• Complete discretization. Seldom interested in discretization of just one continuous attribute (unless there is only one such attribute in a data set).
• Simplest result of discretization. In general, the smaller the size of an attribute's domain after discretization, the simpler the rules that are induced from discretized data. As a result, the knowledge encompassed by such an attribute is more general.
• Consistency. Data sets with continuous attributes are almost always consistent (Lc = 1). When a discretization scheme applied to an attribute's values is "bad", an inconsistent data set may be obtained. When this happens, some valuable information is lost. The level of consistency of the new discretized data should be set as close as possible to the original data's level of consistency.
They proposed a method to measure the performance of an attribute (after discretization) based on class entropy of attribute values. The rationale to use class entropy was that if for some block B $\in$ {A}* the class entropy of block B is zero then there exists a concept C in {d}* such that
B $\subseteq$C, i.e., block B is sufficient to describe (in part or whole) the concept C indicating that the current partitioning is good at least with respect to the block B.
However since whole attribute set is studied the average block entropy of an attribute A is calculated according to the following formula
$$M_{\{A^D\}^*} = \frac{\sum_{B \in \{A^D\}^*} \frac{|B|}{|U|} Ent(B)}{|\{A^D\}^*|}$$

Where $\{A^D\}^*$ is the partition induced by the discretized attribute $\{A^D\}$ .A candidate attribute for which M$\{A^D\}^*$ is maximum is selected as the next attribute for re-discretization. The merit of computing average block entropy for an attribute is that re-compute this measure for an attribute which was last picked for re-discretization. This whole procedure transforms the local discretization of attributes into global discretization of attributes. Another method of global discretization of attributes proposed by [11].

ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 3, Issue 2, March 2014

Clusters were formed by computing an m x m distance matrix between every pair of continuous components of examples in U. The entries in the distance matrix correspond to squared Euclidean distances between data points in i-dimensional space. Originally there were m clusters, each of cardinality one. New clusters were formed by merging two existing clusters that exhibit the most similarity between each other i.e. finding two clusters that are separated by the smallest Euclidean distance. After merging these two similar clusters, similarity (distance) to all the other remaining clusters was computed by Lance and Williams Flexible Method, [15]. Given a cluster a and a new cluster bc to be formed from clusters b and c the distance from bc to a is computed as

$$d_{a(bc)} = d_{bc(a)} = \alpha_b d_{ab} + \alpha_c d_{ac} + \beta d_{bc} + \lambda| d_{ab} - d_{ac} |$$

Where α_b = □α_c = 1/2, β= – 1/4 and λ□ = 0 for the Median Cluster Analysis method.

Clusters formed induced a partition on the set of examples U. The formation of cluster continued until the level of consistency of the partition {K | K I a cluster} was equal to or greater than the original data's level of consistency Lc. This process stops when the condition of level of consistency fails [7][8].

Let r be the number of clusters produced. Let K be a cluster. The set of data points of an attribute A_j (1 ≤ j ≤ i) that define a cluster K is

$$DP_{A_j}^{K} = \{x_j^e | e \in K\},$$

The defining interval IK,Aj of cluster Kj with respect to attribute A_j is

I_{K,A_j} = [L_{K,A_j},R_{K,A_j}] = **[min (**$\mathrm{DP}_{A_j}^{K}$ **), max (**$\mathrm{DP}_{A_j}^{K}$**)]**

For each attribute Aj where j Î {1, 2... i} two sets can be constructed, L_{A_j} and R_{A_j}, which contain the left and right boundary points respectively of the defining intervals I_{K_l,A_j} for l Є {1, 2... r}. Hence,

L_{A_j} = { L_{K_l,A_j} **| l Є {1, 2... r}} and** R_{A_j} = { R_{K_l,A_j} **| l Є {1, 2... r}}.**

This induces partition on the domain of the attribute A_j which is equal to

π_{A_j} = **{[**min_1 **(**L_{A_j}**),** min_2 **(**L_{A_j}**)), [**min_2 **(**L_{A_j}**),** min_3 **(**L_{A_j}**))... [**min_r **(**L_{A_j}**), max (**R_{A_j}**)]},**

Where min_a (L_{A_j}) corresponds to the a^{th} smallest element of L_{A_j}. After the formation of these clusters [11] merged the intervals.

If pAj = {[a0, a1), [a1, a2)... [ak–1, ak] be the interval partition of attribute Aj. Than any pair of adjacent intervals [al-1, al) and [al, al+1) can be used safely into one interval Il–1,l+1 = [al–1, al+1) if and only if the class entropy of the block associated with (Il–1, l+1) 0. A class entropy value of 0 indicates that the new interval (Il–1, l+1) describe one concept only (in part or in full). To prioritize interval merging, they computed the class entropy of the block associated with every pair of adjacent intervals for all continuous attributes picked a pair of adjacent intervals for merging whose class entropy was the smallest amongst continuous attributes. They concluded that Cluster Analysis method had outperformed the Equal Interval Width and Equal Frequency per Interval methods whereas Minimal Class Entropy showed no significant performance difference to any other methods in the global approach to discretization.[9][10][11][12] [9] proposed a 2-step discretization approach to discretize continuous datasets. Discretization was achieved based on discernibility relations of Rough Set Theory. In 2-step discretization approach, instead of taking all the possible basic set of cuts, they had taken the set of cuts returned from MD-heuristics approach. For optimization of the cuts Genetic Algorithm was used to reduce the superfluous cuts among them. They concluded that the entire discretization process was completed in reduced time when compared to the discretization approach described by in [10].

IV. PRESENT WORK

A research design is a master plan that specifies the methods and procedures for collecting and analyzing the needed information. It is a strategy for how the data will be collected and provides the scheme for answering research questions. Research design maintains control to avoid bias that may affect the outcomes.

ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 3, Issue 2, March 2014
The secondary data is obtained from the examination section of the Bharati Vidyapeeth Deemed University, Pune. Masters in Computer Applications (MCA) is a professional program of three years having six semesters. Result repository of MCA program covers the course wise marks of students. Each semester has 7 subjects and thus the total number of subjects including compulsory and elective is about 40. Data mining techniques is applied on this population to answer the above stated research question. The first step of data mining is discretization of data. For the present study, the conditional attributes are the compulsory subjects in MCA program which decides the result of students. The number of such conditional attributes is about 20. Marks obtained by students in these compulsory subjects ranges from 0-100. In order to have the better understandability of the marks scored by students, the range of marks is subdivided into groups, instead of dealing with the whole range of marks.

Two ways of discretization used in present study
Student's marks are the reflection of his studies or his/her knowledge in the subject. University gives the marks out of hundred. These hundred marks are in the ratio of 80:20. University conducts the main exam (i.e. theoretical out of 80), whereas 20 marks form the basis of internal assessment. In the present study we concentrated only on the main exam conducted by the university leaving the internal marks as we are interested only in the study of performance of students in the theoretical exam. Since the marks are out of 80, we converted the obtained marks of student into percentage. In this research a 10-point grading system is used to discretize the range of marks into groups as shown in the following table 1.

Table 1: 10 Point grading system

MARKS	GRADE POINT
[75-100]	10
[70-74.9]	9
[65-69.9]	8
[60-64.9]	6
[55-59.9]	7
[50-54.9]	5.5
[45-49.9]	5.0
[40-44.9]	4.5
[00-39.9]	0

This 10-point scale will be easy to handle the selected data as this will discretize the continuous range of 0-100 marks into 9 discrete intervals.
The result of the student depends on the conditional attributes. The final result of student is declared by evaluating the average marks scored at the completion of the program. These marks contain only the total of theoretical marks out of 80.
The result constitutes the decision attribute. Grades from F (fail) to D (distinction) are used which discretize the result in the manner as shown below in table.2.

Table 2: Discretization of Decision Attribute (Final results)

MARKS	GRADES	EXPLAINATION
[70-100]	O	D, Distinction
[60-69.99]	A+	F, First class
[55-59.99]	B+	HS, Higher Second class
[50-54.99]	B	S, Second class
[40-49.99]	C	P, Pass class
[00-39.99]	F	Fl, Fail class

Using the above criteria, all the subject marks along with the result are discretized.

In the exam conducted by university, student gets marks. These marks are the absolute marks. In the first method of discretization we converted the marks into percentage and then discretized them according to the corresponding

ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 3, Issue 2, March 2014

grades. However these absolute marks have some variations such as the effect of teacher who taught the subject, effect of paper pattern, effect of examiner's evaluation on the students answer sheet, effect of book that student referred etc. These variations can be removed by normalization of the absolute data. This practice is not uncommon as all credit based system uses such normalized data for awarding grades to students.

In the process of normalization, we have calculated the average and the standard deviation for each individual subject. Measures of central tendency (or statistical averages) tell us the point about which items have a tendency to cluster. Such a measure is considered as the most representative figure for the entire mass of data. It refers to the point of greatest concentration of the set of numbers. The mean score represents the average value of all the individual scores in the group. Mean of subject is defined as the value which we get by dividing the total of the numbers of students in given subject by total number of students.

The standard deviation is the most frequently calculated measure of variability or dispersion in a set of data points. The standard deviation value represents the average distance of a set of scores from the mean or average score. This number represents the degree of spread-out-ness of the scores from the mean. The larger is the number, the greater is the spread. It is useful in comparing the variability's in different groups of scores.

$$\sigma = \sqrt{\frac{\Sigma(x-\mu)^2}{n-1}}$$

Where σ = standard deviation
Σ = sum of
x = individual mark of a subject
μ = mean of all marks of a subject
n = sample size (number of marks)

After calculating the average marks and standard deviation of individual subject, we have computed the normalized data by using the following formula:

Normalized data $x' = \frac{x-\mu}{\sigma}$

Discretizing the Normalized data

All the absolute data is converted into the normalized data and hence different approach of discretization is used. In this method, for the present study subject mean and subject standard deviation both are used in making the discrete intervals which are represented in table 3.

Table 3: Intervals based on discretization

Interval	Grade
[min{32, (μ - σ)}, (μ-σ))	F
[(μ-σ), (μ - 0.5σ))	E
[(μ - 0.5σ), μ)	D
[μ, (μ + 0.5σ))	C
[(μ + 0.5σ), (μ + σ))	B
[(μ + σ), max{64, min{(μ + 1.5σ), 72}})	A
[max{64, min{(μ + 1.5σ), 72}}, 80]	O

V. CONCLUSION

It has been observed that discretization is a preprocessing task which when conducted leads to very good results in declaring the rules between attributed, classifications of objects and predicting of classes. How the educational data set behave after discretization can be studied in depth and can be used to predict such similar set of dataset behavior in future..

In this paper two different techniques are used for discretization of the educational data. The first method converts the absolute marks into percentage and a 10 point grading which is used for discretization. In the second method absolute marks are normalized and then with the help of mean and standard deviation discretization of normalized marks are carried out. The second method tries to remove the effects of variations caused due to teaching, paper setting and paper evaluation which purely considers the impacts of the performance of individual students in their respective subjects

ACKNOWLEDGEMENT

The authors sincerely acknowledge the research guidance provided by Dr. M. S. Prasad, Professor, Department of Computer Applications, Bharati Vidyapeeth University's IMED, Pune.

REFERENCES

[1] Bacardit, J. and Garrell J.M., (2004) Analysis and improvements of the Adaptive Discretization Intervals knowledge representation. Genetic and Evolutionary Computation Conference (GECCO04). Lecture Notes in Computer Science 3103, 726-738

[2] Bacardit, J.and Garrell, J.M.. (2003) Evolving multiple discretizations with adaptive intervals for a Pittsburgh Rule-Based Learning Classifier System. Genetic and Evolutionary Computation Conference (GECCO03). Lecture Notes in Computer Science 2724, 1818-1831.

[3] Bertelsen, R., and Martinez, T. R. (1994) Extending ID3 through discretization of continuous inputs. In Proceedings of the 7th Florida Artificial Intelligence Research Symposium (pp. 122–125). Florida AI Research Society

[4] Boulle, M., Khiops (2004) A Statistical Discretization Method of Continuous Attributes. Machine Learning 55:1 53-69

[5] Burdsall, B., and Giraud-Carrier, C. (1997) Evolving fuzzy prototypes for efficient data clustering. In Proceedings of the Second International ICSC Symposium on Fuzzy Logic and Applications (ISFL'97) (pp. 217–223).

[6] Catlett, J. (1991) On changing continuous attributes into ordered discrete attributes. In Proc. Fifth European Working Session on Learning. Berlin: Springer-Verlag, pp. 164–177.

[7] Cerquides, J. and Mantaras, R.L. (1997). Proposal and empirical comparison of a parallelizable distance-based discretization method. In KDD97: Third International Conference on Knowledge Discovery and Data Mining, pp. 139–142.

[8] Chan, C.-C., Batur, C., and Srinivasan, A. 1991. Determination of quantization intervals in rule based model for dynamic. In Proceedings of the IEEE Conference on Systems, Man, and Cybernetics. Harlottes vile, Virginia, pp. 1719–1723.

[9] Chebrolu and Sanjeevi (2012) Rough set theory for discretization based on Boolean reasoning and genetic algorithm, IJCCR, Volume 2, Issue 1, January 2012, pp.22

[10] Chen, C. Y., Li, Z. G., Qiao, S. Y. and Wen, S. P., (2003) "Study on Discretization in Rough Set Based on Genetic Algorithm", Proceeding of the Second International Conference on Machine Learning and Cybernetics, Xi'an, China, pp. 1430-1434,

[11] Chmielewski, M.R. and Grzymala-Busse, J.W. (1994) Global discretization of continuous attributes as preprocessing for machine learning. In Third International Workshop on Rough Sets and Soft Computing, pp. 294–301

[12] Divina, F., Keijzer, M., Marchiori, E (2003) A method for handling numerical attributes in GA based inductive concept learners. In: GECCO 2003: Proceedings of the Genetic and Evolutionary Computation Conference, Springer 898–908

[13] Divina, F., Marchiori, E., (2005) Handling continuous attributes in an evolutionary inductive learner. IEEE Transactions on Evolutionary Computation 9:1 31-43

[14] Dougherty, J.,Kohavi, R., and Sahami, M. (1995) Supervised and unsupervised discretization of continuous features. In Proc. Twelfth International Conference on Machine Learning. Los Altos, CA: Morgan Kaufmann, pp. 194– 202.

[15] Everitt, B. (1980) Cluster Analysis. London, United Kingdom: Heinmann Educational Books, Second Edition,

[16] Fayyad, U. and Irani, K. (1992) On the handling of continuous-valued attributes in decision tree generation. Machine Learning, 8:87–102.

[17] Ferrandiz, S., Boull, M., (2005) Multivariate discretization by recursive supervised bipartition of graph. IV International Conference in Machine Learning and Data Mining in Pattern Recognition (MLDM05). Lecture Notes in Computer Science 3587, 253-264

[18] Giráldez, R., Aguilar, J.S., Riquelme J.C., (2003) Natural Coding: A More Efficient Representation for Evolutionary Learning. Genetic and Evolutionary Computation Conference (GECCO03). Lecture Notes in Computer Science 2723, 979-990

[19] Ho, K.M. and Scott, P.D. (1997). Zeta: A global method for discretization of continuous variables. In KDD97: 3rd International Conference of Knowledge Discovery and Data Mining. Newport Beach, CA, pp. 191– 194.

[20] Holte, R.C. (1993). Very simple classification rules perform well on most commonly used datasets. Machine Learning, 11:63–90

[21] Huan Liu, Farhad Hussain, Chew Lim Tan, Manoranjan Dash Discretization: An Enabling Technique, Data Mining and Knowledge Discovery, 6, 393–423, 2002 c 2002 Kluwer Academic Publishers. Manufactured in The Netherlands.

[22] Kerber, R. (1992) Chimerge: Discretization of numeric attributes. In Proc. AAAI-92, Ninth National Conference Articial Intelligence. AAAI Press/The MIT Press, pp. 123–128

[23] Kotsiantis Sotiris, Dimitris Kanellopoulos (2006) Discretization Techniques: A recent survey, GESTS International Transactions on Computer Science and Engineering, Vol.32 (1), 2006, pp. 47-58

[24] Kwedlo, W., Kretowski, M.: (1999)An evolutionary algorithm using multivariate discretization for decision rule induction. In Proc. of the European Conference on Principles of Data Mining and Knowledge Discovery 392–397

[25] Michal R. Chmielewski and Jerzy W. Grzymala-Busse (1996) Global Discretization of Continuous Attributes as Preprocessing for Machine Learning, Draft. Int. Journal of Approximate Reasoning 15, 319–331

[26] Nguyen, S.H. and Skowron, A. (1995). Quantization of real value attributes: rough set and Boolean reasoning approach. Proceedings of the Second Joint Annual Conference on Information Sciences, pp. 34-37, Wrightsville Beach, NC, September 1995

[27] Nguyen, H.S. (1998). Discretization problem for rough sets methods. Proceedings of the First International Conference on Rough Sets and Current Trends in Computing, pp. 545-552, Warsaw, Poland, June 1998, Springer-Verlag

[28] Shannon, C. and Weaver, W. (1949) The Mathematical Theory of Information. Urbana: University of Illinois Press.

[29] Thornton, C.J. (1992) Techniques of Computational Learning: An Introduction. Chapman and Hall.

[30] Quinlan, J.R. (1986) Induction of decision trees. Machine Learning, 1:81–106.

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

[32] Van de Merckt, (1990) Decision trees in numerical attribute spaces. Machine Learning, 1016–1021

[33] Wang, K. and Liu, B. (1998) Concurrent discretization of multiple attributes. In Pacific-Rim International Conference on AI, pp. 250–259.

[34] Yang, Y., and Webb, G. I. (2001) Proportional k interval discretization for naive-Bayes classifiers. 12th European Conf. on Machine Learning (ECML'01) pp. 564–575.

[35] Z. Pawlak, Z. (1982), Rough sets. International Journal of Computer and Information Sciences 11, pp. 341-356.