

Attribute Selection via a Novel Interval Based Evaluation Algorithm: Applied on Real life data sets

Mostafa A. Salama^{1,a} and Ghada Hassan^{1,b}

¹The British University in Egypt, Cairo, Egypt

Abstract. Real life problems handled by machine learning deals with various forms of values in the data set attributes, like the continuous and discrete form. Discretization is an important step in the pre-processing stage as most of the attribute selection techniques assume the discreteness of the input values. This step could change the internal structure of the input attribute values with respect to the classification problem, and thus the quality of this step directly impact the quality of the selected features. This work discusses the problems existing in the current discretization techniques and proposes an attribute evaluation and selection technique to avoid these problems. Attributes are evaluated in its continuous form directly without biasing its internal structure and enhances the computational complexity by eliminating the discretization step. The basic insight of the proposed approach relies on the inverse relationship between class label distribution overlap and the relative information content of a given attribute. In order to estimate the validity of this assumption, a series of data sets were examined using several standard approaches including our own implementation, and the approaches ranked with respect to the overall classification accuracy. The results, at least with respect to the testing data sets deployed in this study, indicate that the proposed approach outperformed other methods selected for evaluation in this study. These results will be examined over a wider range of continuous attribute data sets from non-medical domains in order to investigate the robustness of these results.

1 Introduction

Attributes in real life data set have various forms that require a pre-processing stage before applying machine learning techniques. The quality of the classification techniques in machine learning and resulted accuracy is limited by the characteristics of the input attributes [1, 2]. The first characteristic is the type of the input attributes, discrete, ordinal, and continuous attributes[3]. Then the second aspect is the relevance of the input attributes to the targeted classification problem. Some attributes may contain irrelevant or redundant attributes that do not contribute to effective classification. Irrelevant attributes could have a negative effect on the resulted classification accuracy and its existence in the data set increases the effort and complexity in the machine learning technique applied [4, 5]. Attribute reduction technique is applied to select only the relevant attributes to the target class labels. Attribute selection algorithm is applied on two stages, First the attribute evaluation followed by the subset attribute subset selection. An attribute evaluation function ranks the attributes and removes the attributes that do not achieve an adequate score, while subset selection searches the set of possible attributes for an optimal subset. Most of the attribute evaluation techniques as statistical techniques have to be preceded by a discretization method. Few type of

algorithms are in-dependent on the attribute selection and discretization process like for instance, genetic algorithm (GA), as it may be applied as a attribute selection wrapper to search for an optimal attribute subset for dimensionality reduction [7] [8]. But due to the high number combinations of attributes (subsets) these methods are computationally expensive in many real life cases. Some approaches integrate the attribute selection method directly into the classification pipeline. Techniques like ChiMerge attribute selection method utilize an internal discretization approach. Another method [9] tries to find the attribute ranks without discretization but this method is dependent on Information gain, which is only applicable for nominal data.

The main problem that lies in Discretization algorithms is its alteration of the internal structure of the attribute values. This can weakened the interdependency between the attribute values and the target class labels. And thus it can potentially influence subsequent classification results [6]. Discretization algorithms are either unsupervised or supervised algorithms. Un-supervised discretization is performed independently of the attribute selection and the classification tasks and performs an equal frequency or width discretization. Supervised discretization like information-gain, entropy-based, statistics based algorithms that are based on measuring the number class specific values for each attribute. Another technique like Class-Attribute Dependent Discretizer (CADD) measures

^ae-mail: mostafa.salama@bue.edu.eg

^be-mail: ghada.hassan@bue.edu.eg

the dependency between each class and each interval of attribute values. Other than the discretization step is considered an extra step in the Data mining process that increases its time complexity, these techniques are considered of a high complexity cost.

This study proposes a supervised attribute evaluation technique that can be applied directly on all forms of attribute values. The purpose of this technique is to avoid the use of the discretization method in the pre-processing stage to decrease the computational complexity and to preserve the information quality. The algorithm of the attribute evaluation here is based on evaluating attributes according to number of class specific values or range of values. In other words, as the intervals of sorted attribute values related to a specific class increases as the importance of this attribute increases. The high lengths of the non-overlapped intervals for each attribute show the ability of this attribute on discriminating the classification of the instance. The discrimination power of this attribute is an evidence for the important of this attribute.

This algorithm operates on the premises that attributes are sorted, and then the number of adjacent values that have a single class label (not overlapped with other classes) is calculated. It quantizes the entire space of the class-dependent feature values into non-overlapping intervals. The resulted intervals generated by a different method from the Vector quantization method. The algorithm in this work search for the highest length regions that could exist in-between each adjacent group of points. While vector quantization divides a large set of points into groups of adjacent points; each group is represented by a centroid point as in k-means algorithm. It models the probability density function of the points' distribution. The length of the region in-between each two adjacent groups is not considered in vector quantization. This approach set the default number of intervals is three, but this number may not be appropriate in some cases (data sets). Thus, the algorithm is repeated for up-to 10 intervals until the maximum classification accuracy is reached.

The approach adopted here (termed interval based attribute selection) will be compared with standard discretization based approaches such as IB, Relief Feature, ChiMerge and Information Gain. The comparison criteria will be applied by evaluating which attribute selection method yields maximal classification accuracy with the smallest number of attributes. More than ten benchmark data sets will be used in the comparison study, and the same classification techniques will be used after all the applied pre-processing techniques for fairness guarantee. The classification methods used here is Naive Bayesian Tree method, it is considered as an evaluation function for the forward attribute selection method. This method is implemented in the Weka suite of machine learning tools. The results from this study indicate that the approach adopted in this technique is superior to traditional based discretization methods in terms of the classification accuracy of standard classifiers.

The rest of this paper is organized as follows. Section 2 is a literature review on the problems existing in different attribute selection techniques and discretization tech-

niques. Section 3 shows the proposed interval based attribute selection method. Classification results and comparisons with different attribute selection methods are illustrated in section 4. Lastly, Conclusion is discussed in section 5.

2 Literature review

Pre-processing stage include two independent steps which are the attribute discretization step and the attribute evaluation and selection step. Most of the attribute selection and machine learning techniques are performing on data in a discrete form. Several problems appears in the discretization techniques that have a direct negative effect on the classification accuracy results. First, two famous attribute evaluation techniques are demonstrated to show the dependence on the values to be in a discrete form. Then a brief discussion on the current discretization techniques and the existing problems in these techniques is provided.

2.1 Attribute evaluation techniques

The chi-Square χ^2 method measures the lack of independence between each attribute and the target class. ChiMerge or Chi2-Square is a χ^2 -based discretization method [10] [11]. It uses the χ^2 statistic to discretize numeric attributes repeatedly until some inconsistencies are found in the data. Adjacent intervals with the least χ^2 values are merged together, because low χ^2 values for a pair indicates similar class distributions. This merging process proceeds recursively until all χ^2 values of all pairs exceeds a parameter *signlevel* (initially 0.5). Then repeat the previous steps with a decreasing *signlevel* until an inconsistency rate is exceeded, where two patterns are the same but classified into different categories.

The mutual information *MI* (also called cross entropy or information gain) is a widely used information theoretic measure of the stochastic dependency of discrete random variables [12] [13]. The mutual information $I(A; C)$ between values of attribute *A* and the set of classes *C* can be considered as a measure of the amount of knowledge of *C* provided by *A* (or conversely on the amount of knowledge of *A* provided by *C*). In other words $I(A; C)$ measures the interdependence between *A* and *C* as $H(C) - H(C|A)$, where The entropy $H(C)$ measures the degree of uncertainty entailed by the set of classes *C*, and The conditional entropy $H(C|A)$ measures the degree of uncertainty entailed by the set of classes *C* given the set of attribute values *A*.

2.2 Discretization techniques

The discreteness of the input attributes is an important factor in many of the classification techniques. The nature of real life data usually has continuous spectrum of data values in the attribute vectors. Discretization method is applied by setting a set of cut points to partition the range of values into a small number of intervals. The term cut point is a real value inside a range of values

that divides this range into two separate intervals. The algorithm of the discretization method has to preserve both, the information and statistical quality of the data. The information quality ensures that the internal structure of the continuous values according to the present classification problem is preserved. While the statistical quality guarantees that the number intervals is the minimum number and data values generalization is sufficient for the investigated problem. Discretization methods are either non-supervised or supervised by the target class label. The non-supervised methods divide the data values into equal range of values or into equal number of values. The number of intervals in these methods is a user defined threshold that is based on the domain problem. Example of the non-supervised methods is based on the k-mean clustering algorithm. This method randomly assigns k data values in the continuous range of values to be centers of the k intervals. The rest of the values in this range are assigned to the intervals of the nearest corresponding k-values. On the other hand, the supervised methods divide the data values based on the target class labels. The Entropy-based and chi-square based algorithms are examples of supervised partitioning. Entropy-based partitioning calculates the entropy for all the points in a range of values and selects the split point that has the lowest entropy value. The entropy value is based on the probability of the instances on both sides of the points are belonging to each class label. The ChiSquare partitioning merges two adjacent intervals of values into one interval if these intervals are independent on the target class label. Both of these two supervised discretization methods are binary splitting the continuous spectrum of data values until a splitting criterion is satisfied. The splitting criteria are dependent on user defined value based on the domain problem. A refinement stage in the discretization method is the test of the generated splits through an evaluation method. Classifiers test are applied as evaluation were the point of the minimum error rate is selected as the split point. Another way to make use of the evaluation tests is using of different discretization method and selecting the best fitting method for the domain data attributes [14, 15]. One general problem in non-supervised and supervised discretization method is the user interference during the process. The user interference could identify the number of intervals in non-supervised discretization or identify a threshold as a stopping criterion in supervised discretization. Although a default value can used for general cases, this shows inaccurate classification results in machine learning. The independence of the unsupervised techniques on the target class label leads to the loose of the internal structure of the attribute values and decreases the information quality of the data. The resulted intervals after applying the discretization method may contain values that are corresponding to different classes. This leads to ignoring several data values that could be helpful in classification accuracy results in machine learning. The interference of the target class labels in the discretization process is important to define the correct boundaries between the different data value. On the other

hand, supervised methods suffers problems specific to each method. For example, the chimerge algorithm tests each adjacent interval only before deciding the interval merge and ignores the other surrounding intervals. This could leads to the formation of large intervals that do not correctly represent the correct class-based structure of the data distribution. Another example is the information entropy algorithm that tests each value in all the existing attributes, and iteratively does this test several times until the stopping condition is reached. On the other hand, Class-Attribute Dependent Discretizer (CADD) measures the estimated joint probability of the event that a value belongs to a particular class and a particular range of values. These algorithms are relatively expensive, as number of values of attributes in real life problems is very high. Another general problem is that the addition of extra step in the pre-processing stage, which is discretization, increases the effort in the machine learning process. It is required to decrease the number of steps in the machine learning techniques in order to easily identify the problem behind the classification accuracy error.

3 Interval based attribute selection technique

3.1 Discrimination concept

The interval based (I.B.) attribute selection technique developed in this work is an attribute evaluation techniques that does not require a preceding discretization technique. The technique evaluates each attribute separately and uses the target class labels for evaluating the discrimination power of this attribute. The algorithm of this technique works for discrete and continuous values equally, as it considers the range of the values in the attribute even if this range contains a single value. The proposed technique is based on a concept that is implemented by supervised attribute discretization and evaluation algorithms as shown here in sections 2. Theoretically, this concept states that if a specific attribute value appears for only a single target class label, then this value discriminates the class of the containing instance. And as the discriminating value increases in an attribute, then the value of this attribute increases. Discretization methods group the adjacent discriminating continuous values together in a single partition or interval. And also group the adjacent class independent values in a single interval. This concept is implemented in various techniques and proved practically in different areas. The proposed technique uses this concept in an evaluation technique that is applied directly on the continuous values, and avoids the drawbacks of the currently evaluation techniques. Accordingly, the applied algorithm in this technique sorts the values in each attribute and detects the intervals where the values appears in a single class and the intervals that contains and intersection between classes. As the intersection between attribute value ranges of different class labels decreases as the importance of this attribute increases.

The rationale for this approach is that if an attribute has a certain continuous range of values which appears only in the case of a specific class label, then this attribute is necessary as it provides required information for the given class label. Further, the length of the range of values increases in proportion to the information content, this content is in terms of the classification accuracy. In order to clarify the applied concept in this technique, figure 1 shows the range of sorted values of a single attribute. This example assumes the existence of three classes in the problem, and the range of values corresponding to each class label is as shown in the figure. In this figure, the dashed areas show the ranges of values that are not overlapped between multiple classes; only a single class label is assigned to this label. As the length of this dashed area increases as the importance of this attribute increases. These non-overlapped intervals are summed, and the rank of the attribute is the resulted sum relative to the total interval of values of this attribute.

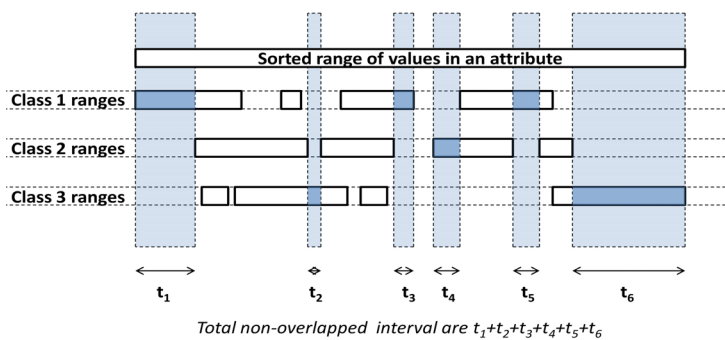


Figure 1. The non-overlapped intervals of an attribute categorized into three classes example

Equation 1 shows mathematically the evaluation of an attribute a based on calculated non-overlapped as shown in figure 1. In this equation, the μ value represents the evaluation of the attributes and it is measured for every attribute in the input data set.

$$\mu_a = \frac{1}{\max_a - \min_a} * \sum_{i \in I_n} t_i \quad (1)$$

μ_a represents the rank of attribute a ;
 \max_a is the maximum value of the attribute a ;
 \min_a is the minimum value of the attribute a ;
 I_n is set of non-overlapped intervals;
 t_i is the length of non-overlapped interval i in I_n ;
 Theoretically, as the number of discriminating values in the attribute increases, as the attribute importance to the classification problem increases. It is clear in equation 1 that μ value is directly proportional to number of values lies in each class solely. The division by the total length of values of the attribute is for the normalization of the resulted value. This step standardizes the attribute rank to guarantee the fairness in the comparison to the other attributes ranks.

3.2 Algorithm implemented

Algorithm 1 formalizes the steps of evaluating a single attribute a by calculating μ_a in equation 1. This algorithm is applied on every attribute in the data set, and the attributes of the highest μ is selected based on the forward feature selection algorithm. The number of the intervals in an attribute is referred to as d . This number is directly proportional to the number of values in the attribute. The factor of proportionality is a user defined value. The number of attribute' intervals of the example in figure 1 is four. For simplicity, the value d is an input in algorithm 1.

Assume the number of intervals for each class is d
 μ_a : Attribute a 's rank, initial value is 0
 l_a : length of range of values in attribute a , i.e. $\max_a - \min_a$
 x_a and n_a : max and min values of attribute a
for Each Class label c **do**
 Sort values of attribute a corresponding to class c in an ascending order
 n_c : number of values in attribute a corresponding to class c , a_c
 Create array $D[n_c]$: array of difference between each value a_c and the previous value in the sorted list of values of attribute a
 Search and locate the maximum d values in array $D[n_c]$
 Create a two dimensional array $R_c[d, 2]$, each of d pairs of values $[s, e]$
 Where s and e are the located values containing the maximum value in array $D[n_c]$
end for
 Sort values of attribute a in an ascending order
 $t = 0$: non-overlapped interval
for Each Class label c **do**
 for Each interval $[s, e]$ in array $R_c[d, 2]$ **do**
 Find interval $[s_i, e_i]$ that has no intersection with other intervals in the rest of the class labels
 $t = t + (e_i - s_i)$
 end for
end for
 $\mu_a = t/l$
 return μ_a
Algorithm 1: Calculate Interval based rank μ_a of attribute a

The removal of misleading values in Algorithm 2 is an optional step as the accuracy depends on the collection methodologies. This step decreases the sensitivity to outliers by removing the values that are farthest away from the average value of the attribute values. This step should remove only a small percentage of the values in the attribute in order to ensure maximization of the overall classification accuracy.

Finally, the proposed technique sorts the attributes according to their μ value in a descending order, and select the best attributes using forward feature selection algorithm.

Input : *Interval* values of an attribute *a* for objects lies in class *c*

Output : *avg* average of the values of an attribute *a* in a class *c*

for *x* * *IntervalLength* values **do**

Remove the value of max difference from the average *avg* .

end for

Algorithm 2: Remove percentage *x* of misleading

3.3 Comparison to other attributes evaluation methods

It is well established that the classification accuracy increases as the number of selected attributes increases up to a point [2]. In many instances, including additional attributes may actually reduce the overall classification accuracy [17]. This behavior is depicted in figure 2.

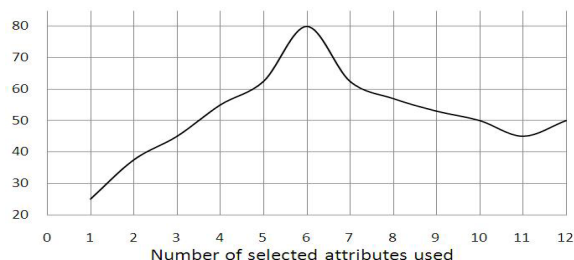


Figure 2. Classification accuracy versus the number of selected attributes

One explanation for the lack of a linear dependence of attribute number and classification accuracy is that certain attributes possess conflicting information with respect to the decision classes. Further, when the attributes of highest rank (relevance) are used the classification accuracy increases while the correlated or non-informative attributes can behave like noise in the data, thereby degrading classifier performance. In this work, the behavior of the attribute selection approach is examined in order to determine whether this non-linear relationship between attribute cardinality and classification accuracy exists across a set of three attribute selection methods. It should be noted that from an operational perspective, the best attribute selection method should adhere to the following criteria:

- The peak (maximum accuracy) is reached with the smallest number of attribute.
- The peak is of the highest value among other attribute evaluation methods.

The comparison will be validated by multiple training and testing runs. Further, initially, the data set will contain only the most relevant attribute (as determined empirically through the various comparative approaches), followed by the inclusion of the other lower ranked attributes until all have been included. This method could be considered as a semi-wrapper method as the evaluation will be applied only on a certain subset of attributes. The wrapper-based

approaches employ an induction classifier as a black box using cross-validation or bootstrap techniques. Other related approaches include the deployment of genetic algorithm to evaluate the attribute subset candidates [8]. This approach has issues with respect to the high computational costs, and also has the added difficulty of not dealing with data sets containing continuous attributes. Therefore, in this work, three different classifiers will be applied: a Naive Bayesian Tree, a support vector machine and multi-layer perceptron (both via Weka implementations). Both have been extensively used as classification tools with a great deal of success from object recognition. The attribute selection methods used are Chi-merge, gain ration, and information gain attribute selection methods.

4 Experimental results and discussion

4.1 Practical Evaluation of the Interval Based Feature Evaluation

Two criteria will be used to evaluate the proposed interval based feature evaluation technique. The first criterion is the classification accuracy percentage of the selected attributes. The second criterion, in case the same classification accuracy is resulted from the usage two different feature evaluation technique, the number of selected features is considered as a discrimination factor. The feature evaluation technique that leads to the lowest number of features is considered better than the others. Seven different benchmark data sets are used to compare the proposed technique to three other different techniques. The data sets are representing real life problems from different domains. Each contains a number of continuous valued attributes which are typically discretized when analyzed [16]. Further, these data sets are realistic in other ways, in that they typically do not possess specific distribution (at least one is not required typically for analysis) of values and may contain misleading and/or missing values, due to an error in calibrations or collection of data. The used data sets are obtained from the UCI machine learning repository [18] which are parkinsons, sonar, ringnorm, bupa, Hebatitis, NSLKDD, Thrombosis, wdbc, and ionosphere data sets. The number of features and the number of instances in each data set are as following respectively, parkinsons 96 22, sonar 208, 60, ringnorm 7400, 20, bupa 288, 6, Hebatitis 66, 19, wdbc 424, 30, NSLKDD 2000, 15, and Ionosphere 250, 34. The used feature evaluation techniques used in this experimental study are the ReliefF, the ChiMerge and the InfoGain techniques. Forward feature selection algorithm is used for the selection of the lowest number of feature that leads to the highest classification accuracy percentage. Finally, Naive Bayesian Tree classification method is used as the evaluation function. The classification accuracy was calculated by dividing the number of correctly classified objects divided by the total number of objects in the testing data set. 10-fold training and testing of the input data set method is used to generate a consistent classification accuracy percentage results. Also, in order to ensure the

Table 1. Classification accuracy and number of features comparison, $d=3$

Data set	IBFE	ReliefF	ChiMrg	InfoGain
parkinsons	94.79	94.79	93.75	92.70
	19	20	2	3
sonar	81.73	81.25	80.28	79.80
	25	6	14	17
ringnorm	97.70	97.62	97.51	97.22
	20	19	19	19
bupa	67.36	62.5	56.25	56.25
	4	4	5	5
Hebatitis	66.66	65.15	65.15	65.15
	1	1	1	1
NSLKDD	90.2	89.55	89.60	89.35
	13	14	13	15
Thrombosis	44.11	44.11	44.11	44.11
	1	2	2	2
wdbc	95.04	96.46	94.81	94.81
	12	4	16	16
ionosphere	90.4	91.2	89.2	89.6
	21	7	2	3

fairness of the test, instances are divided equally into two class labels. An important note to be discussed later in this practical evaluation of the model is number of intervals. The number of intervals for all the data sets used in table 1 is three intervals only. This table shows for each data set the classification accuracy after feature selection, and the number of selected features below. For example, when using the IBFE feature selection technique on the parkinsons data set, the classification accuracy is 94.79% and the number of selected features are 19.

For some data sets like HS_AS_MR and HS_AR_MS in [19], the usage of three intervals only may be too small to lead to high classification accuracy percentages. In case of using six intervals $d = 6$ in algorithm 1 for the HS_AS_MR and HS_AR_MS data sets, the classification accuracy shows a higher percentages than in case of using three intervals only and than the other 'compared to' feature evaluation techniques. Figure 3 shows how the classification accuracy and number of selected features varies according to the used number of intervals. This figure shows that the lowest number of feature and the highest classification accuracy is reached when the number of intervals is 6 intervals.

Also in case of an increasing value d for the indian-diabetes data sets from [18], the classification accuracy shows a similar output like other feature evaluation techniques, but the number of features decreased from 8 to 7 features only. The number of features and the number of instances of data sets are HS_AS_MR 100, 74, HS_AR_MS 100, 74 and indian-diabetes 8, 536. Table 2 shows the results of the classification accuracy and the corresponding number of features # that leads to the highest classification accuracy.

Table 2. Classification accuracy and number of features comparison, $d>3$

Data set	IBFE	ReliefF	ChiMrg	InfoGain
HS_AS_MR	94.59	93.24	93.24	91.89
	13	4	11	5
HS_AR_MS	86.48	86.48	83.78	86.48
	69	8	15	7
Indian-diabetes	77.79	77.79	77.79	77.79
	7	8	8	8

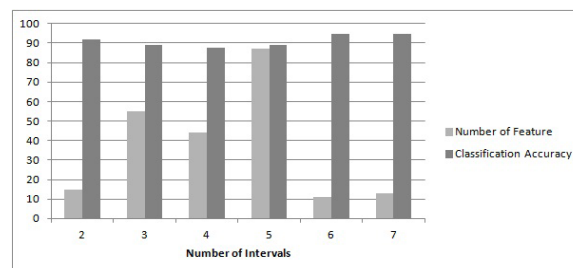


Figure 3. The variation of the classification accuracy and number of selected features according to the number of intervals

Finally, another test is applied for other classification techniques, like SVM for Support Vector Machine, mlp for Mult-Layer Perceptron and j48 for Decision Tree. The resulted classification accuracy for svm and mlp is the same as the other feature evaluation technique. But it is noticed that the classification accuracy of j48 after selecting features using the proposed evaluation technique is lower than the ReliefF, chimerge and InfoGain techniques. After increasing the number of intervals, the accuracy percentage after applying the proposed techniques is higher than the other techniques. The reason of such behavior can be justified based on the characteristics of the Decision Tree. The input attributes to the Decision Tree should be in a discretized form. As the discreteness of the values in the attribute increases, as the number of intervals should increases. And Dealing with the attribute values as continuous using the interval based evaluation method could decrease the quality of the Decision Tree classifier.

5 Conclusion

Most of the attribute evaluation techniques depend on the assumption the input data set is in a discrete form. The application of discretization method preceding the evaluation technique is crucial in these cases. If the discretization method is avoided, the alteration of the internal structure resulted from this method and the extra pre-processing are avoided. The results from this study support this general claim echoed loudly across the data mining community at large. The approach developed in this work relies on calculating the number of adjacent values that are class specific. The approach was compared with known attribute reduction methods and classifiers. The performance was evaluated with respect to the classification accuracy and as a function of the number of selected attributes. The approach was applied to a set of 12 publicly available data

sets from different domains. The results demonstrate that the proposed attribute selection scheme was superior to the other approaches, regardless of the classifier deployed. After applying the proposed feature evaluation selection, the classification accuracy percentages are high and the numbers of feature are low in comparison the usage of the other feature evaluation techniques. These results were clearly present across all the twelve data sets. This provides a wide range of very diverse data sets that contain a variety of levels of sparseness and mixtures of ordinal, discrete, and continuous data sets.

References

- [1] Michael K. Ismail, Vic Ciesielski, An Empirical Investigation of the Impact of Discretization on Common Data Distributions, Proceedings of the Third International Conference on Hybrid Intelligent Systems (HIS 03), pp. 692-701, 2003.
- [2] Andreas G. K. Janecek and Wilfried N. Gansterer and Michael A. Demel and Gerhard F. Ecker, On the Relationship Between Feature Selection and Classification Accuracy, JMLR: Workshop and Conference Proceedings, vol.4, pp. 90–105, 2008.
- [3] Manuel Mejia-Lavalle, Guillermo Rodriguez, Feature Selection with Interactions for Continuous Attributes and Discrete Class, Proceeding CERMA 07 Proceedings of the Electronics, Robotics and Automotive Mechanics Conference, pp. 318-323, 2007.
- [4] Mostafa A. Salama and Aboul Ella Hassanien and Aly A. Fahmy, Pattern-based Subspace Classification Model, Proceedings of The Second World Congress on Nature and Biologically Inspired Computing, IEEE Xplore Digital Library, Kitakyushu, pp. 357-362, Dec. 2010.
- [5] Mithun Nagendra Prasad and Arcot Sowmya and Inge Koch, Designing Relevant Features for Continuous Data Sets Using ICA, International Journal of Computational Intelligence and Applications, vol. 7, pp. 447-468, 2008.
- [6] Huan Liu and Rudy Setiono, Feature Selection via Discretization, IEEE Transactions on Knowledge and Data Engineering, vol. 9, pp. 642-645, Aug. 1997.
- [7] Khan, A., and Revett, K., Data mining the PIMA dataset using rough set theory with a special emphasis on rule reduction, bookIEEE International Multitopic Conference (INMIC 2004), Islamabad, Pakistan, pp. 334-338, 2004.
- [8] Chi-ho Tsang and Sam Kwong and Hanli Wang, Genetic-fuzzy rule mining approach and evaluation of feature selection techniques for anomaly intrusion detection, Pattern Recognition, vol. 40, pp. 2373-2391, 2007.
- [9] Manuel Mejda-lavalle and Eduardo F. Morales and Gustavo Arroyo, Two Simple and Effective Feature Selection Methods for Continuous Attributes with Discrete Multi-class, bookMexican International Conference on Artificial Intelligence, pp. 452-461, 2007.
- [10] Huan Liu and Rudy Setiono, Chi2: Feature Selection and Discretization of Numeric Attributes, bookInternational Conference on Tools with Artificial Intelligence, Virginia, USA, pp. 388-392, 1995.
- [11] Abdelwadood Moh'd, Chi Square attribute Extraction Based Svms Arabic Language Text Categorization System, Journal of Computer Science, vol.3, pp. 430-435, 2007
- [12] Hanchuan Peng and Fuhui Long and Chris Ding, Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy, bookIEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27, pp. 1226-1238, 2005.
- [13] Jin-Jie Huang and Yun-Ze Cai and Xiao-Ming Xu, A parameterless attribute ranking algorithm based on MI, bookNeurocomputing, vol.71, pp. 1656-1668, 2008.
- [14] Sotiris Kotsiantis, Dimitris Kanellopoulos, Discretization Techniques: A recent survey, GESTS International Transactions on Computer Science and Engineering, Vol.32 (1), pp. 47-58, 2006.
- [15] Rajashree Dash, Rajib Lochan Paramguru, Rasmita Dash, Comparative Analysis of Supervised and Un-supervised Discretization Techniques, International Journal of Advances in Science and Technology, Vol. 2, No. 3, 2011.
- [16] K. Revett, M.S. Szczuka, P. Jahankhani, and V. Kodogiannis, Attribute Selection for EEG Signal Classification Using Rough Sets and Neural Networks, bookThe 5th International Conference Rough Sets and Current Trends in Computing (RSCTC 2006), Kobe, Japan, pp. 408-417, 2006.
- [17] Mostafa A. Salama, Abol Allah Hassanien, Aly E. Fahmy, Pattern-based Subspace Classification Model, Proceeding of the second World Congress on Nature and Biologically Inspired Computing (NaBIC), Kitakyushu, Japan, pp. 357-362, Dec. 2010.
- [18] UCI, Machine Learning Repository, <http://archive.ics.uci.edu/ml/datasets.html>.
- [19] Maglogiannis I, Loukis E, Zaropoulos E and Stasis A., Support vectors machine-based identification of heart valve diseases using heart sounds. Comput. Methods Programs Biomed, pp. 47-61, Mar 6. 2009 .
- [20] Tamilarasan, A., S. Mukkamala, A.H. Sung and K. Yendrapalli, Feature Ranking and Selection for Intrusion Detection Using Artificial Neural Networks and Statistical Methods, Proceedings of the International Joint Conference on Neural Networks, pp. 4754-4761, 2006.