A VPRS based Approach for Enhancement in the Classification Efficiency of Machinery Datasets

Rashmi Upadhyay
Department of Computer Science,SIMS

Ritesh Shah
Department Of Computer Science,SIMS

ABSTRACT
In this paper an effective approach is proposed to enhance the classification accuracy of machinery datasets. Since in the today's world of data estimation, data is increasing at explosive rate. At the same time that bulk of data may contain relevant or irrelevant data. Presence of rough data in the dataset's creates hindrance in doing classification of huge amount of data. In this paper a clear picture is depicted of removing noisy and rough data from the datasets. We have selected a wide range of machinery data from UCI dataset repository for our study. Several learning a technique from different paradigms leads to the output extracted to be independent of the underlying classifier. In this paper, a novel classification technique called VPRS is used so to relax the subset operator. We have also used Levenberg-Marquardt algorithm to classify and compare the outcome results of training and testing phase. First of all, test attribute space is optimized and the attributes which are not correlated with the decision attributes are deleted. Experiments proved that the accuracy of dataset's have got increased by using the combination of K-fold technique. Firstly the data sets are applied to VPRS algorithm which reduces useless attribute and these reduced attributes will get applied to Levenberg - Marquardt classifier to classify the datasets. This process is repeated for 5 times by using K-Fold Technique, and finally we have removed rough data from the machinery data set and increased classification accuracy.

Keyword
K-fold technique, VPRS, ANN, Livenberg Marquardt Algorithm.

1. INTRODUCTION
Classification of machinery datasets is considered to be one of the important tools of data mining to turn large amount of "passive data" into useful "actionable information". The main idea of classification is taking the information theory as a tool and selecting the important attribute; classify the datasets until certain condition is satisfied. Classification accuracy may be increased as a result of feature selection through the removal of noisy, irrelevant, or redundant features. For data sets which are smaller in size, the runtimes of learning algorithms can be improved significantly. Classification is the prediction approach in data mining techniques. There are many algorithms based on classification that is Instance Based, Bayesian Networks, Support Vector Machine and Decision tree, Neural Networks. Artificial Neural Networks [1] are a family of information processing techniques inspired by the way biological nervous systems process information. The fundamental concept of neural network is the structure of the information processing system composed of a large number of highly interconnected processing elements or neurons, a neural network uses the human-like technique of learning by example to solve problems. A Neural network is often configured for a specific application, such as data classification or pattern recognition, through a learning process called training. Just as in biological systems, learning involves adjustments to the synaptic connections that exist between the Neural networks are being applied to an increasingly large number of real-world problems. Their primary advantage is that they can solve problems that are too complex for conventional technologies, problems that do not have an algorithmic solution, or for which an algorithmic solution is too complex to be defined. Data-mining algorithms have been employed in many classification problems. In this paper, a hybrid idea of classification method has been adapted by using k-fold technique as a partitioning method of datasets to improve accuracy of classification. The experiments done on the datasets, show that the pre-classification can improve the accuracy of classification. In this paper, a novel technique of increasing efficiency of the classified machinery datasets is used. The attribute reduction based on knowledge dependence is used to optimize the test attribute space of datasets by using VPRS[3]. The reduced data set has the same classification capability as the original set. After this, knowledge dependence based on rough set theory on condition attribute, is used as heuristic information for selecting test attributes from the reduction set. Larger the degree of dependency of attributes, more certain information is included in it. Levenberg-Marquardt classifier classifies the datasets repeatedly by using k-fold cross validation technique. K-fold cross validation technique devides the datasets in train data and test data. The datasets are classified repeatedly until consistent output comes. Experiments proved that VPRSANN using K-Fold technique has reduces the complexity and of the datasets and increase the classification accuracy in every fold's. The result of the example shows that this technique overcomes the deficiency of the rough datasets and provides superior classification techniques. Rest of this paper is organized as background details are discussed in section 2. In section 3 proposed approach is described. In section 4 performance evaluation and some comparisons are given. Then the concluding remarks are given in section 5.

2. BACKGROUND
Rough set theory (RST), proposed by Poland mathematician Pawlak in 1982, is a new mathematic tool to deal with vagueness and uncertainty. Its main idea is that to classify samples into similar classes containing objects that are indiscernible with respect to some attributes. RST [9] can solve many problems occurred in data reduction, feature selection and pattern extraction so that we can get rid of redundant data even in the information system with null values or missing data. Rough set theory is a mathematical technique which is used to analyze imprecise, uncertain, or vague information. Knowledge reduction and knowledge dependence are the central content of rough set theory. Researchers have developed many variations and proposed methods for the classification of rough datasets that use rough set theory such as the attribute core, reduct, and discernibility matrix to improve the classification accuracy of machinery datasets.

However, these proposed approaches also have their limitations. They only do well in accurate classification where objects are strictly classified according to equivalence classes; hence the induced classifiers lack the ability to tolerate possible noises in real world datasets. In order to improve the shortcomings of rough set model, the classical rough set model is extended, Ziarko proposed a variable precision rough set model , which introduced the β(0≤β<0.5) [1] based on the basic rough set model, and allowed some degree of misclassification rate. Aijun also proposed a variable precision rough set model, which introduced β(0.5≤β<1)) as the correct rate. The main concept of variable precision rough set theory is degree of dependency and significance of attributes which is used in the proposed algorithm to select splitting attribute, therefore this approach proposes a new attribute selection criterion, the enhanced information gain based on degree of β-dependency and significance of condition attributes on decision attribute is used as a heuristic for selecting the optimal splitting attribute to overcome problem of attribute reduction and also extends variable precision rough set theory.

2.1 VPRS

In data analysis, Variable Precision Rough Set (VPRS) is very useful for addressing problems where data sets have lots of boundary objects. VPRS attempts to introduce an element of "fuzziness" to the rough set model and hence (although indirectly) utilize the boundary region information. The principal idea behind VPRS is to allow objects to be classified with an error smaller than a certain (manually) predefined level. However, the introduction of this threshold is contrary to the rough set ideology of operating only on the information contained within the data itself [1] [6]. We have introduced some basic concept of VPRS theory [2].

2.1.1 Information Systems

An information system [5] is a pair S= (U, W, V, f) where U is a non-empty finite set of objects called universe. W denotes the set of attributes, it is usually divided into two subsets P and Q, which denote the set of condition attributes and the set of decision attribute, respectively. f:U×W→V is an information function, where V = a∈WVa is the domain of attribute.

***A.* Information Systems**

An Variable precision rough sets (VPRS) [4] attempts to improve upon rough set theory by relaxing the subset operator. It was proposed to analyze and identify data patterns which represent statistical trends rather than functional. The main idea of VPRS is to allow objects to be classified with an error smaller than a certain predefined level. This approach is arguably easiest to be understood within the framework of classification. Let P, Q ⊆ U, the relative classification error is defined by C(P,Q) = $\begin{cases} 1 - \frac{|P \cap Q|}{|P|} & |P| \\ 0 \end{cases}$

Where |P| is the cardinality of that set

***B.* Degree of inclusion**

Let P, Q be any two sets, if 0 ≤β< 0.5 , the majority inclusion relation can be defined as:

P⊆ β if C (P, Q) ≤ β , 0 ≤β< 0.5

***C.* β -lower and β - upper Approximation of Set**

Let R be the indiscernible relation on the universe U. Suppose (U, R) is an approximation space.U/R = {P1, P2…Pn} where Pi is an equivalence class of R. For any subset P⊆ U, lower approximation R^{β}-P and upper approximation R-$_{\beta}$P of P with precision level β respect to R is respectively defined as

$$R^{\beta} P = U\{ Q \in U/R \left| \frac{P \cap Q}{P} \leq \beta \right. \}$$

$$\overline{R}_{\beta} P = U\{Q \in U/R | \frac{P \cap Q}{P} < 1\}$$

Where the domain of β is 0 ≤β<0.5, Rβ-X is also called β-Positive region (POS (P, Q)). The β boundary of P with respect to R is defined as:

BNDβ P = U{Q∈U/R|β<(P∩Q)/P+<1-β}

When β= 0, Ziarko variable precision rough set model becomes Pawlak rough set model.[6]

3. PROPOSED MODEL

In this section basic preliminaries behind ANN and then the Levenberg Marquardt Approach is discussed.

3.1 ANN

ANNs [8] are models inspired in biological neural networks whose signal-processing-topology is composed of several simple-processing-elements named neurons or nodes, which are highly interconnected among them and operates in parallel system as represented in Fig.1. A Neural network is often configured for a specific application, such as data classification or pattern recognition, through a learning process called training. Just as in biological systems, learning involves adjustments to the synaptic connections that exist between the neurons. The fundamental concept of neural networks is the structure of the information processing system. Composed of a large number of highly interconnected processing elements or neurons, a neural network uses the human-like technique of learning by example to solve problems. Just as in biological systems, learning involves adjustments to the synaptic connections that exist between the neurons. Neural networks can differ based on the way their neurons are connected, the specific kinds of computations their neurons perform and the way they transmit patterns of activity throughout the network. Neural networks are being applied to an increasingly large number of real-world problems. Their primary advantage is that they can solve problems that are too complex for conventional technologies, problems that do not have an algorithmic solution, or for which an algorithmic solution is too complex to be defined. [3].

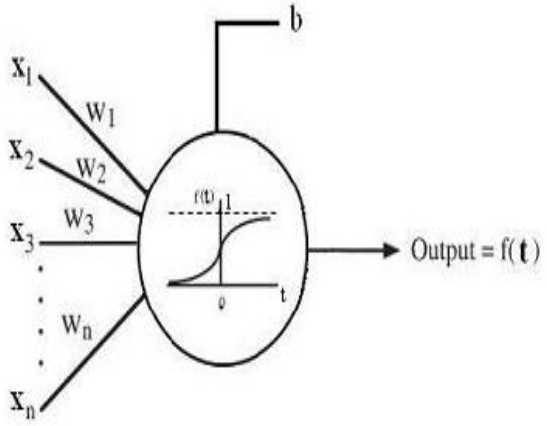

Fig. 1. Representation Of ANN

3.2 Levenberg- Marquardt Back Propogation Algorithm

A number of learning rules are available to train neural networks. The Back Propagation (BP) learning algorithm another technique which is widely used to train multi-layer feed-forward neural network in parallel systems. A simple diagram of back propagation algorithm is shown below where signals are received at the input layer, pass through the hidden layer, and reach to the output layer, and then fed to the input layer again for learning. The learning process primarily involves the determining of connection weights and patterns of connections. The BP neural network approximates the non-linear relationship between the input and the output by adjusting the weight values internally instead of giving the function expression explicitly.

The BP algorithm uses gradient descent method to look for the minimum value of the estimated error function. Evaluated minimum error function which is estimated by using combination of weights of the input data will be consider as solution of the learning problem. Levenberg-Marquardt algorithm is another algorithm based on back propagation algorithm. Levenberg-Marquardt algorithm [4]was designed to approach second-order training speed without having to compute the Hessian matrix. The LM algorithm is an iterative technique that locates a local minimum of a multi-variate function that is expressed as the sum of squares of several non-linear, real-valued functions. It has become a standard Technique for non- linear least-squares problems, widely adopted in various disciplines for dealing with data-fitting applications. LM can be thought of as a combination of steepest descent and the Gauss-Newton method. When the current solution is far from a local minimum, the algorithm behaves like a steepest descent method: slow, but guaranteed to converge. When the current solution is close to a local minimum, it becomes a Gauss-Newton method and exhibits fast convergence. When the performance function has the form of a sum of squares (as is typical in training feed forward networks), then the Hessian matrix can be approximated as H=JT J and the gradient can be computed as g=JT e where J is the Jacobian matrix that contains first derivatives of the network errors with respect to the weights and biases, and e is a vector of network errors. The Jacobian matrix can be computed through a standard back propagation technique (see [HaMe94]) that is much less complex than computing the Hessian matrix. The Levenberg-Marquardt algorithm uses this approximation to the Hessian matrix in the following Newton-like update:

xK+1=xk-[JT+μI]-1JTe

When the scalar μ is zero, this is just Newton's method, using the approximate Hessian matrix. When μ is large, this becomes gradient descent with a small step size. Newton's method is faster and more accurate near an error minimum, so the aim is to shift toward Newton's method as quickly as possible. Thus, μ is decreased after each successful step (reduction in performance function) and is increased only when a tentative step would increase the performance function. In this way, the performance function is always reduced at each iteration of the algorithm.

3.3 K-Fold Cross Validation Technique

K-fold Cross validation technique[7] is one of the technique which is widely used to improve upon residual evaluation model. Residual evaluation model was not able to solve the problem of over fitting while K-fold technique[10] gave the best solution to overcome the problem of over fitting-fold technique does not give entire datasets when training as a learner. The main concept used behind cross validation was to separate the datasets in k subset so to get train data and test data. During separation some data is removed from training set, Thus when training is done, the data which was removed will be used as test data and can be used to test the performance of the learned model.

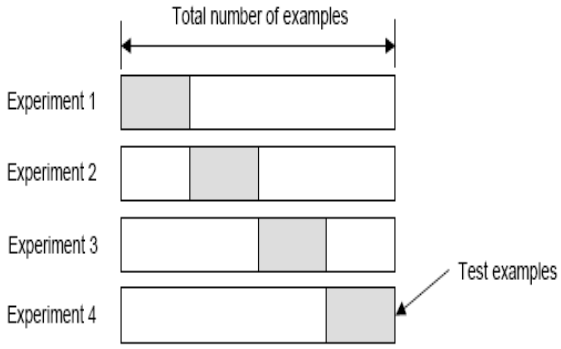

Fig. 2. Representation Of K-Fold

Function approximator plays one of the important role during training of the datasets. Function approximator use fits a function on training sets. During testing Function approximator predicts the output values for the data and accumulates the error to find out the mean absolute error. This process is repeated for each subset. Each time, one of the k subsets is used as the test set and the other k-1 subsets are put together to form a training set. The main advantage of this method is every data in the datasets are tested exactly once and gets to be in a training set k-1 times.

$$E = \frac{1}{K}\sum_{i=1}^{K} E\,i$$

4. EXPERIMENTS AND RESULTS

Our experiments are carried out on an Intel (R) Pentium(R) CPU B940@ 2.00 GHz, 2GB RAM, 32 bit Windows 7 Operating System. All procedures were implemented on MATLAB System. We have use four groups of datasets from the UCI Machine Learning Repository. In the experiments self test validation was conducted on all data sets to calculate the classification accuracy .The performance chart shows that the classifier accuracy has increased from 94.66 % in 2-fold to 98.33% in 5-fold. First of all we have focuses on dividing the machinery data sets in multiple-folds so to partitioned the datasets into train data and test data. K-fold cross-validation technique partitioned, the original sample randomly k subsamples. Of the k subsamples, a single subsample is retained as the validation data for testing the classifier, and the remaining k − 1 subsamples are used as training data. The cross-validation process is then repeated k times, with each of the k subsamples used exactly once as the test data. The k results from the folds are then averaged to produce a single performance estimation. To divide the applied datasets into training and testing parts Random Division technique is used. This technique separate targets into three sets: training, validation, and testing using random indices.

[trainInd , valInd , testInd] = dividerand (Q, trainRatio ,valRatio, testRatio)

It takes the following inputs,

- Q :Number of targets to divide up

- trainRatio: Ratio of vectors for training.

  Default = 0.7

- valRatio: Ratio of vectors for validation.

  Default = 0.15.

- testRatio: Ratio of vectors for testing.

  Default = 0.15.

And returns

- trainInd — Training indices
- valInd — Validation indices
- testInd — Test indices.

After the division of datasets the corresponding Test data and train data are classified separately, and then classifier performance is evaluated. During training phase VPRS will be applied to the training data so that data dependency between different attribute will get calculated. The main advantage of VPRS technique is that it allows for a controlled degree of misclassification in its formalism which, in turn, leads to more general notions of set approximations. Thus data dependencies in situations where data items are considered independent by the original rough sets model will get achieved. The criteria of checking beta dependency using VPRS is shown below.

```
k=length(pos1)/length(dec_data)
if (k<1&k>0)
   msgbox('partial dependence ')
elseif k==0
    msgbox('not dependence ')
else
    msgbox('fully dependence ')
end
```

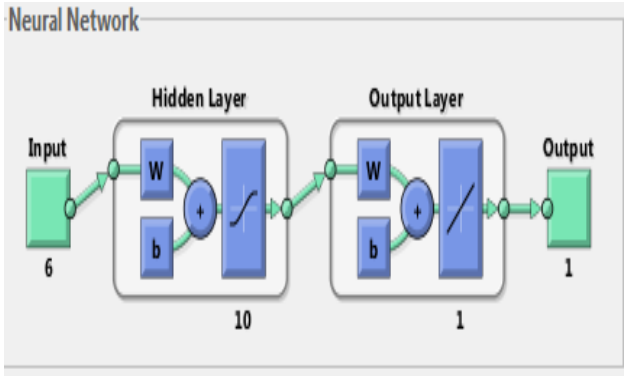

Fig. 3. Representation Of Neural Network

Under this Rough feature selection technique were the dependency of every attribute is calculated and if suppose any irrelevant attribute is recognized then that attribute will get removed as it is rough data and thus it will remove unnecessary attributes which may affect both rule comprehension and rule prediction performance. After this these datasets will get applied on Levenberg- Marquardt Back Propagation Algorithm. Levenberg-Marquardt algorithm trains the datasets to obtained the desired classified datasets. During traing network, training function is used which updates the weight and bias values according to Levenberg-Marquardt optimization. This is often the fastest back propagation algorithm , and is highly recommended as a first-choice supervised algorithm, although it does require more memory than other algorithms. The training process is shown below where trainlm is the training function.

Trainlm (net,TR, trainV, valV, testV) takes these inputs)

Where,

Net	Neural network
TR	Initial training record created by train
trainV	Training data created by train
valV	Validation data created by train
TestV	test data created by train

And returns

net	Trained network
TR	Training record of various values over each epoch.

Finally the datasets will get trained and classified accordingly. And then the result of test data and train data will be compared. Now holdout method is repeated 5 time's. Each time one of the *k* subsets is used as Test data set and the other k-1 subsets are put together to form the New training one. Finally, the k accuracy estimates are averaged to provide the accuracy of the final model trained on new training dataset's. The variance term of the resulting estimate reduces as *k* increases. After completion of 5-fold technique no more training samples will be there to classify, finally the datasets will be classified with more information gain. Performance of evaluation Of classified machinery heart datasets are shown below. Result shows that at 2-fold state the performance of classifier which was later increased to 91 % at 5-fold. Thus efficiency of datasets has got increased from an average of 80% (test data) to 91 %(train data). The performance graph of the result output is shown below. In the figure we have shown the performance analysis and classification result of Heart data sets.

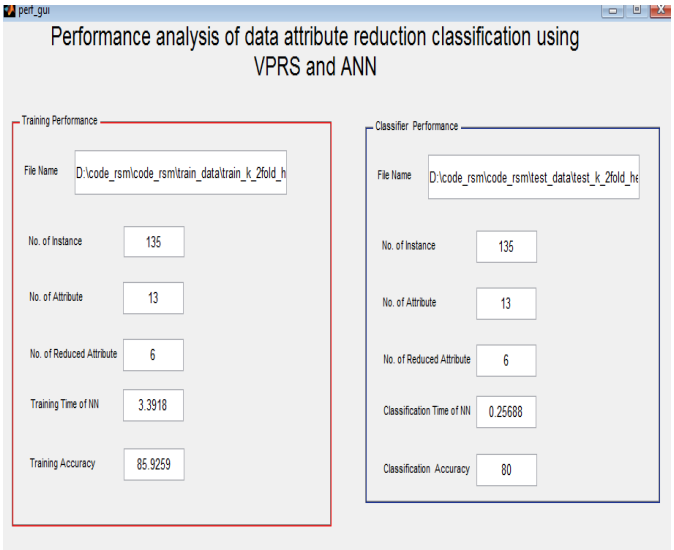

Fig. 4. Representation of Classification Result at 2-Fold

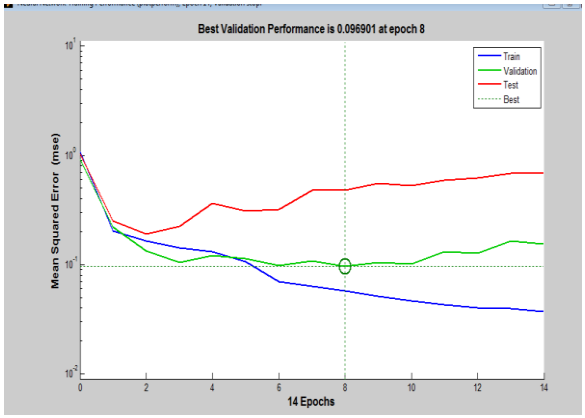

Fig. 5. Representation of Best Validation Performance at 2-Fold

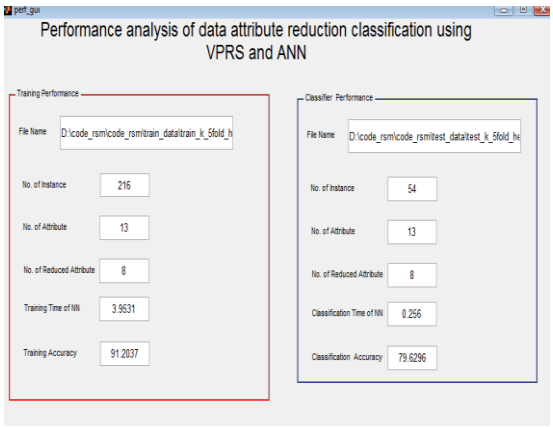

Fig. 6. Representation of Classification Result at 5-Fold

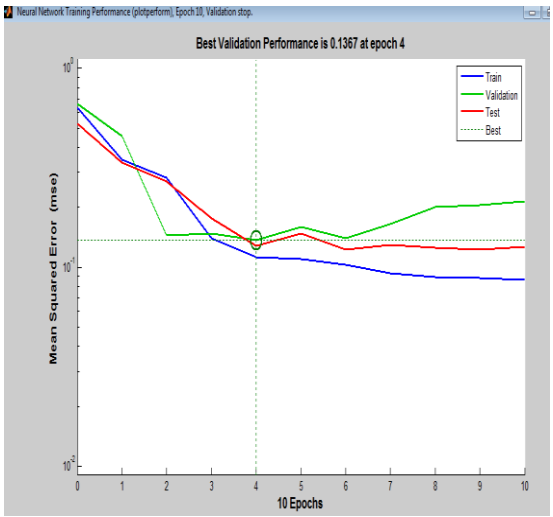

Fig. 7. Representation of Best Validation Performance at 5-Fold

5. CONCLUSION

We proposed the concept of the enhanced information gain based on VPRS Model. The analysis of the results has been carried out by means of the proper statistical study, which shows the goodness of this approach for dealing with classified machinery dataset's this approach improved the classification rate and proposes new attribute selection criteria. K-Fold cross validation technique has evaluated the performance of classifier. A new technique is proposed dataset shift has potentially introduced which result in inaccurate performance estimation. This paper analyzes the prevalence and impact of partition in the field of classification. This model has produced stable performance in the era of classification. Thus Datasets are properly classified

6. REFERENCES

[1] Khaled Shabam, "A Cascade of Artificial Neural Networks to Predict Transformers Oil Parameters," Prediction approach, vol.16,pp.517-518,2009.

[2] Rajkumar Sharma, "An Optimize Decision Tree Algorithm Based on Variable Precision Rough Set Theory Using Degree of β -dependency and Significance of Attributes ,"Machine Learning, vol. 2,pp.3942-3947,2012 .

[3] Wojciech Ziarko, "Variable Precision Rough Set Model," Classification Approach,pp.40-41,1993.

[4] Bahram G. Kermani, "Performance of the Levenberg–Marquardt neural network training method in electronic nose applications" Machine learning,vol.110,pp.15-16,2005.

[5] Z. Pawlak, "Information systems theoretical foundations," Information Systems, vol. 6, pp. 205–218, 1981.

[6] Manolis I.A.," Is Levenberg-Marquardt the Most Efficient Optimization Algorithm for Implementing Bundle Adjustment?"Optimization algorithm,pp. 1-3.

[7] Jose García Moreno-Torres,"Study on the Impact of Partition-Induced Dataset Shift on k-fold Cross-Validation,"vol.23,pp.1305-1306,2012.

[8] Djaffar Ould Abdeslam," A unified artificial neural network archtecture for active power filters", vol.54,pp.61-62,2007.

[9] Suyun Zhao,"The model of fuzzy varaible precesion rough sets",vol.17,pp.451-452,2009.

[10] Jose García Moreno-Torres,"Study on impact of partition-induced dataset shift on k-fold cross validation",vol.23,pp.1304,2012.