Biased Confidence Classification Algorithm for Faculty Subject Allocation in Education Domain

Abhishek Kumar Agrawal[1], Akshay Gupta[2], Venkatesan M.[3]

School of Computing Science and Engineering, VIT University
Vellore, INDIA

[1]abhishek.agrawal15@yahoo.com
[2]akshay3004@gmail.com
[3]mvenkatesan@vit.ac.in

Abstract— **There are certain selection and allocation processes in the real world that are performed based on previous knowledge. We can find many applications related to these processes, for example, Candidate selection for promotion, Candidate to department allotment, faculty to subject allotment, etc. In the small scale, the process is not very tedious. But, when it comes to large scale selections and allotments, the process can be very time consuming and prone to human error. So automation in this process is the need of the hour. This process requiresan ample amount of decision making, and so data-mining techniques can prove to be effective methods to deal with such problems. There are many multi-class classification methods that can be used as the solution to these problems. But, decisions based only on trained classifiers with historical data-patterns, won't be sufficient in the real time allotment. There might be certain parameters that should be given more priority for the current allocation. In this paper we combine both the historical trends and biased parameters to perform the classification satisfying real time demands. We then compare and contrast its performance against various existing classification algorithms. Our experiment with faculty-course allotment dataset shows that this method is more suitable than other methods for such practical applications.**

***Keywords*— Classification, Cross-Validation, Kernel Methods**

I. INTRODUCTION

Since the inception of data-mining techniques [17], they have been implemented in various decision making processes. Still, there are many practical applications where these techniques can make a remarkable difference as compared to traditional approaches. In the domain of selection and allotment process [19], these techniques can be a major success. Certain real time applications like candidate selection for promotion, employee department allotment, faculty subject allotment, etc. while performing on a large scale can be very tedious and time consuming.

In such application, where different parameters, qualities and performance related to subjects are compared, the results obtained by manual allotment approach may be prone to error. Hence the automation in such process can overcome the possibility of errors and also reduce the human effort needed for the same. Automation can be achieved using various classification techniques of data-mining. But for the practical applications mentioned above, applying classification algorithms like Support Vector Machine (SVM) [2], Decision Tree [11] and Naive Bayesian Algorithm [3] may not satisfy the real time demand since those classifiers are trained against the historical data-pattern and cannot be biased. This type of decision making process involves real time constraints as well as historical results.

In section II, we give a brief description of these real time constraints along with the meta-data of the datasets which gives us a more precise idea of the complexity of such processes. In section III, we propose an algorithm named Biased Confidence Classification Algorithm, able to deal with such practical application datasets more efficiently. The algorithm provides the option to bias the parameters of the dataset according to the real-time requirement. We discuss the use and importance of kernel methods for the calculation of the biasing functions in Section IV. The results of prediction and cross validation of various classification techniquesare reviewed and compared with Biased Confidence Classification Algorithm in section V.Finally, we have discussions and conclusions in section VI.

II. DATA-SET DESCRIPTION

Allotment process in real time is very complex, and the decision cannot be just made on the basis of classification results on the dataset. In order to have a better understanding of this process, we have taken a dataset for faculty subject allocation process. Here,we have data which describes attributes like teaching experience (in years), level of seniority as designation, area of interests, subject choices, semester, year and subjects taught as a label attribute. So this dataset contains all the faculty subject allotment records through the years. All the attributes have been represented as numerical to provide ease in applying any mathematical model or function [15]. In this type of process, the allotment cannot be just based on the historical pattern. In this application, for each faculty the following parameters can play an important role for allocation process:

(i) No. of times any subject is taught ,

(ii) Most recently taught subjects,

(iii) Area of interests.

Apart from these, subject-choices, faculty-load etc. are the factors that can also be considered for allotment process. The subset of the faculty-subject dataset is shown below :

TABLE I

FACULTY-SUBJECT ALLOTMENT DATA

Faculty code	Designation	Experience	AOI_1	AOI_2	Year	Subject cluster_id
AJU	5	6	5	5	2006	4
MAI	6	5	8	10	2009	8
TAK	2	16	1	1	2007	4
DPA	3	13	6	10	2006	6
HSE	4	16	8	10	2008	8
PKR	2	16	2	6	2010	6
KKU	4	9	7	2	2009	7
RMJ	3	14	7	1	2008	1
MNM	5	10	5	4	2010	4

In the table above, Faculty code represents the unique id for each faculty member, Designation specifies their level of seniority (smaller the value, more the seniority), Experience denotes the Teaching Experience in years, AOI_1 and AOI_2 is symbolic for their areas of interest in teaching. Year and subject cluster_id denotes the subject allotted form the cluster in that year. For example: Subjects related to Computer Networks and Security are mapped to cluster_id 1;similarly Database Management Systems related subjects are mapped to cluster_id 8 and so on. For this application, the real time requirements could be:

(i) More consideration can be given to the subject recently taught,

(ii) Sometimes area of interests can be given preference over subject choices and recently taught subjects,

(iii) Ideal allotment where allotment is done purely based on historical pattern can also be a good choice.

Hence along with the historical pattern, the real time requirements play a vital role in this process.

III. BIASED CONFIDENCE MODEL

The model describes an effective algorithm of supervised learning classification for these types of practical applications. Since the algorithm is application oriented, hence the description involves the terminologies of the application domain. In this case, we have taken faculty subject allotment process.

A. *Allotment Process*

*1) Data Storage:*Every faculty has a history of subjects undertaken. We define a 2-Dimensional array to store the unique subjects with their count which indicates the no. of times that particular subject has been taught by that faculty.

$$X(i, j) = \{ \begin{array}{l} X(i, 0) = s_i \\ X(i, 1) = c_i \} \end{array}$$

Here s_i denotes the i^{th} unique subject, c_i denotes the count of the i^{th} subject till the year previous to the allotment.

2) Confidence Calculation: Confidence of each subject is calculated for each faculty.

$$X(i)' = (c_i / n) . w_i$$

$X(i)'$ denotes the subject confidence**.**

n is the total no. of years taken into account

w_i is the biased weight factor for the i^{th} subject (Biased weight factor will be calculated separately based on the area of interests and recently allotted subjects).

3) Maximum Confidence: The subject having maximum confidence will be allotted to that faculty.

$$Y = \max (X(0)' , X(1)' , X(2)', .., X(i)', ….)$$

Y denotes the subject allotted to the faculty.

This process is repeated for every faculty individually.

B. *Biased Weight Factor*

In Section III-A, we have described weight biased factor in order to evaluate the confidence for each subject. This factor is the key term of this algorithm.. This factor will help in adjusting the allotment error. We can tune this factor in order to obtain higher accuracy in prediction.

$w_i$can be calculated in the following ways:

```
If( X(i,0) == A1 || X(i,0) == A2) then   x=k1
   else x = k2
```

```
If ($X(i,0) == S_n$ || $X(i,0) == S_{n-1}$) then y=k3
  else y=k4

  $w_i=f(x,y)$
```

Here A1 denotes the area of interest 1,
Similarly A2 denotes the area of interest 2,
x is the area of interest parameter,
S_n is the most recently allotted subject to the faculty,
S_{n-1} is the second most recently allotted subject to the faculty,
yis the recent subject parameter.
k1,k2,k3 and k4 are real positive constants.
w_i is the function of x and y. In order to map x and y, to evaluate distinct weight factors in the feature space, we have used different kernel functions. Kernel functions Implementation and results are discussed in section IV.

C. *Accuracy*

In order to test the accuracy of the model, we compute the percentage accuracy.

$$\% \text{ Accuracy} = \frac{\text{No.of true Predictions}}{\text{Total no.of predictions}}$$

Here true prediction can be verified by comparing it to the corresponding year subject allotment.

IV. KERNEL METHODS

After setting the weight factor parameters (x,y), we move for the calculation of biased weight factor. Kernel methods [4,5] are a class of algorithms for pattern analysis, whose best known element is the support vector machine (SVM). The general task of pattern analysis is to find and study general types of relations(for example, clusters, rankings, principal components, correlations, classifications, etc.) in general types of data (such as sequences, text documents, sets of points, vectors, images, etc).
KMs approach [5] the problem by mapping the data into a high dimensional feature space, where eachcoordinate corresponds to one feature of the data items, effectively transforming the data into a set of points in a Euclidean space. In that space, a variety of methods can be used to find relations in the data. Since the mapping can be quite general (not necessarily linear, for example), the relations found in this way are accordingly very general. This mapping approach is called the kernel trick.

A. *Choosing the right Kernel*

Choosing the most appropriate kernel highly depends on the problem at hand - and fine tuning its parameters can easily become a tedious and cumbersome task. The choice of a Kernel depends on the problem at hand because it depends on what we are trying to model. A polynomial kernel, for example, allows us to model feature conjunctions up to the order of the polynomial. Radial basis functions[13,14] allows to pick out circles (or hyper-spheres) - in contrast with the Linear kernel, which allows only to pick out lines (or hyper-planes).The motivation behind the choice of a particular kernel can be very intuitive and straightforward depending on what kind of information we are expecting to extract about the data.

B. *Kernel Functions*

We have applied certain kernel function to compare the classification results.

1) Linear Kernel Function

The Linear kernel [14,7] is the simplest kernel function. It is given by the inner product <x,y> plus an optional constant **c**. Kernel algorithms using a linear kernel are often equivalent to their non-kernel counterparts, i.e. KPCA with linear kernel is the same as standard PCA.

$$k(x,y) = x^T y + c$$

2) *Polynomial Kernel Function*

The Polynomial kernel [14,6] is a non-stationary kernel. Polynomial kernels are well suited for problems where all the training data is normalized.

$$k(x.y) = (ax^T y + c)^d$$

3) Gaussian Kernel Function

The adjustable parameter **sigma**[14**]** plays a major role in the performance of the kernel, and should be carefully tuned to the problem at hand. If overestimated, the exponential will behave almost linearly and the higher-dimensional projection will start to lose its non-linear power.

$$k(x,y) = \exp(-\gamma||x-y||^2)$$

4) Multi-quadratic Kernel Function

This kernel function[14,7] is less computationally intensive than the Gaussian kernel and can be used as an alternative when using the Gaussian becomes too expensive.

$$k(x,y) = \sqrt{||x-y||^2 + c^2}$$

C. *Grid Search Performance*

In order to find the suitable kernel functions, we have performed Grid Search [18] to determine the best set of Biased Factor Parameters of the faculty subject dataset. Here we have linearly varied biased weight factors(x,y) in a range of [0-10] for the best results.

1) *Linear Kernel Function:*

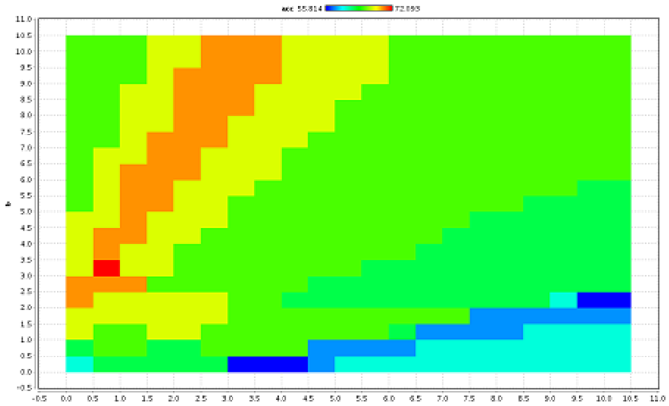

Fig 1: Accuracy Graph using Linear Kernel Function

The linear kernel function has marked up to the highest accuracy but this accuracy is obtained at very few points.

2) *Polynomial Kernel Function:*

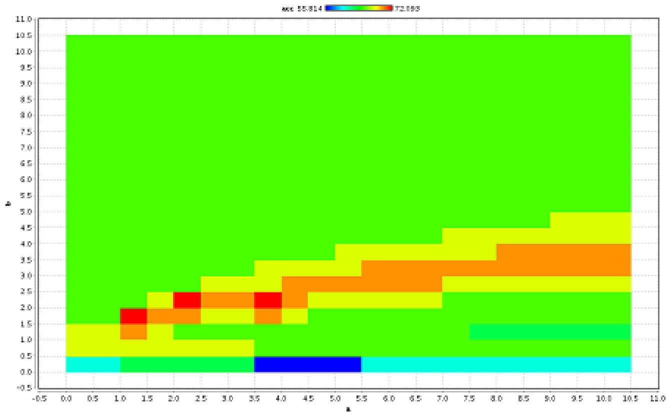

Fig 2: Accuracy Graph using Polynomial Kernel Function

The polynomial Kernel function[14,6] also shows maximum accuracy at relatively less set of points.

3) *Gaussian Kernel Function*

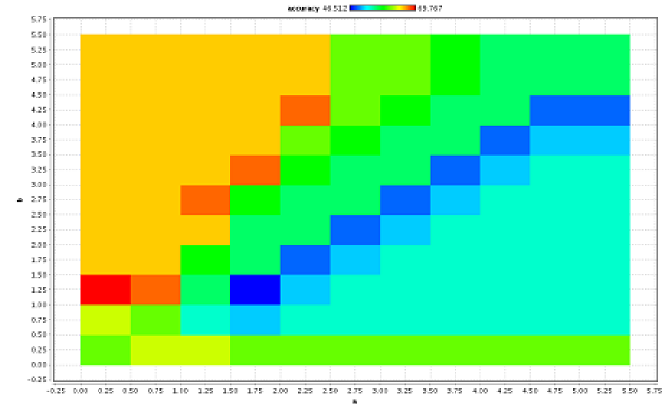

Fig 3: Accuracy Graph using Gaussian Kernel Function

Gaussian Kernel function has maximum accuracyat a number of tested points.

4) *Multi-quadratic Kernel Function*

This is so far the best kernel function to show the maximum accuracy at maximum no. of points.

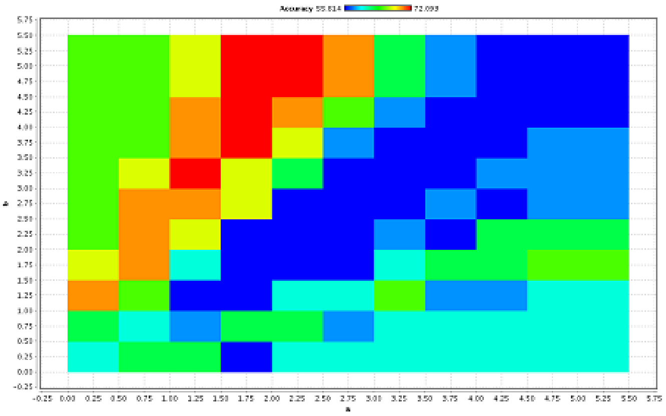

Fig 4: Accuracy Graph using Multi-quadratic Kernel Function

Based on the results obtained from the kernel functions graphs, we have obtained some conclusions which are projected in the table below:

TABLE II

KERNEL FUNCTIONS PERFORMANCE

Kernel Function	X	y	Min Accuracy	Max Accuracy
Linear	1	3.5	55.81%	72.09%
Polynomial	1.5	2	52.81%	72.09%
Gaussian	0.5	1.5	46.51%	69.77%
Multiquadratic	1.5	3.5	55.81%	72.09%

Using the Kernel methods and finding the maximum accuracy actually reveals the most preferred parameters. In the above cases higher value of ‘y’ (recent subject parameter) is preferred over ‘x’(area of interest parameter). Hence for the future prediction this knowledge can be kept in mind.

V. Comparison with Different Classifiers

There exist many classification algorithms that prove to be efficient with data-sets of different applications [1]. We have reviewed the performance of different popular classifiers with the faculty-subject dataset and compared the result with Biased Confidence Classifier.

A. *Cross Validation Report*

In *k*-fold cross-validation, [7] the original sample is randomly partitioned into *k* subsamples. Of the k subsamples, a single subsample is retained as the validation data for testing the model, and the remaining $k-1$ sub samples are used as training data. The cross-validation process is then repeated *k* times (the *folds*), with each of the *k* subsamples used exactly once as the validation data. The *k* results from the folds then can be averaged (or otherwise combined) to produce a single estimation.10-fold cross-validation is the most commonly used method [8], so we have used this method to observe the average performance results of each classifier.

We have reviewed Naive Bayesian Kernel , Rule Induction, Support Vector Machine(SVM), Decision Tree and Biased Confidence classification algorithms for the cross validation.

1) Naïve Bayesian Kernel: 10-fold Cross validation.

The Naive Bayesian Classifier [10,3] technique is based on the so-called Bayesian theorem and is particularly suited when the dimensionality of the inputs is high. The 'kernel' distribution [10] is appropriate for features that have a continuous distribution. It does not require a strong assumption such as a normal distribution and you can use it in cases where the distribution of a feature may be skewed or have multiple peaks or modes. It requires more computing time and more memory than the normal distribution. In the table below, we have reviewed its performance with faculty-subject dataset.

TABLE III
NAÏVE BAYESIAN KERNEL CLASSIFIER

Estimation mode	Min Bandwidth	No. of Kernels	Kappa	% Accuracy
Greedy	0.1	1	0.407	50.30%
Greedy	0.1	5	0.386	48.42%
Greedy	0.5	1	0.413	50.76%
Greedy	0.5	10	0.444	53.07%
Greedy	1	1	0.409	50.28%
Greedy	1	5	0.424	51.23%
Greedy	1	10	0.436	52.19%
Full	NA	NA	0.43	52.64%

Tuning the parameters of Naïve Bayesian Kernel classifier, we have obtained the above results.Maximum of the cross validation accuracy is 53.07%.

2) Rule Induction: 10-fold Cross validation

Rule Induction classifier[16] learns a pruned set of rules with respect to the information gain. Starting with the less prevalent classes, the algorithm iteratively grows and prunes rules until there are no positive examples left or the error rate is greater than 50%.In the growing phase, for each rule, conditions are added greedily to the rule until the rule is perfect (i.e. 100% accurate). The procedure tries every possible value for each attribute and selects the condition with highest information gain.In the prune phase, for each rule any final sequences of the antecedents is pruned with the pruning metric p/(p+n).In the table below, we have reviewed its performance with faculty-subject dataset.

TABLE IV
RULE INDUCTION CLASSIFIER

Criterion	Sample Ratio	Pureness	Kappa	Accuracy
Info Gain	0.9	0.9	0.401	51.15%
Info Gain	1	0.9	0.415	51.23%
Info Gain	0.8	0.9	0.386	49.83%
Info Gain	0.9	0.8	0.339	46.08%
Accuracy	0.9	0.9	0.376	49.65%
Accuracy	1	0.9	0.487	57.60%
Accuracy	0.8	0.9	0.444	54.44%
Accuracy	0.9	0.8	0.473	57.16%

Similarly tuning the parameters of Rule Induction classifier has resulted with the maximum accuracy 57.60% in cross validation. The result is better than the Naive Bayesian classifier.

3) Support Vector Machine:10-fold Cross Validation

In the table below, we have reviewed SVM [2]performance with faculty-subject dataset.

TABLE V
SUPPORT VECTOR MACHINE CLASSIFIER

Kernel	Gamma	Cost	Epsilon	Kappa	Accuracy
RBF	0.125	8	0.001	0.488	57.27%
RBF	0.1	125	0.01	0.483	56.82%
RBF	0.5	512	0.001	0.511	59.57%
RBF	0.125	2048	0.001	0.484	56.84%
RBF	0.1	2048	0.01	0.536	59.77%
RBF	0.125	512	0.01	0.478	56.36%
RBF	0.5	1024	0.001	0.511	59.57%
RBF	0.5	1024	0.001	0.511	59.57%

A support vector machine constructs a hyper-plane or set of hyper-planes in a high- or infinite- dimensional space, which can be used for classification, regression, or other tasks. Intuitively, a good separation is achieved by the hyper-plane that has the largest distance to the nearest training data points of any class (so-called functional margin), since in general the larger the margin the lower the generalization error of the classifier. In most of the data-sets it shows maximum accuracy among the pile of classifiers. Here also among the classifier taken to test the faculty-subject data-set,SVM showed maximum accuracy using RBF(Radial Basis Function)[12].

4) Decision Tree: 10-fold Cross validation

In Decision Tree Algorithm [11] the goal is to create a model that predicts the value of a target variable based on several input variables.

TABLE VI
DECISION TREE CLASSIFIER

Criterion	Min Leaf Size	Confidence	Kappa	Accuracy
Gain Ratio	2	0.25	0.49	57.64%
Gain Ratio	2	0.5	0.48	57.19%
Gain Ratio	1	0.25	0.48	56.67%
Gain Ratio	3	0.25	0.35	47.88%
Info Gain	2	0.25	0.47	56.26%
Info Gain	2	0.5	0.46	55.80%
Info Gain	1	0.25	0.46	55.28%
Info Gain	1	0.5	0.45	54.83%
Info Gain	3	0.25	0.35	47.88%

Decision tree is also one of the most tested and popular classifier. For the faculty-subject data-set,the maximum accuracy obtained through cross-validation is slightly lesser than that of SVM.

So far the results obtained were less and comparable. Hence the results produced so far with these classifiers are tempting for the inception of new algorithm for such application to fill the gap.

5) Biased Confidence : 10-fold Cross validation

Biased Confidence Classification Algorithm as proposed performs classification on certain set of applications where biasing of attributes plays an important role in achieving higher accuracy. The cross validation gives the idea of the weightage of biasing factors in obtaining maximum accuracy.

In the table below, we have reviewed its performance on faculty-subject dataset by tuning biased weight factors(x,y). The cross validation percentage accuracy supports the value of this algorithm.

TABLE VII
BIASED CONFIDENCE CLASSIFIER

Criterion	x	y	Kappa	Accuracy
Accuracy	0.5	5.5	0.617	60.29%
Accuracy	1.5	6	0.606	59.86%
Accuracy	1	5.5	0.625	61.29%
Accuracy	0.9	0.8	0.589	58.43%
Accuracy	1	0.5	0.663	63.20%
Accuracy	3	1.5	0.687	63.56%
Accuracy	0.5	3	0.637	62.70%
Accuracy	1.5	1	0.629	61.82%

B. *Comparative study*

After reviewing the cross validation results, we have summarized our report to get the best suitable algorithm for such applications.

TABLE VII
COMPARISON OF CLASSIFIERS

Classifier	Kappa	Accuracy
Naïve Bayesian	0.444	53.07%
Rule Induction	0.487	57.60%
Support Vector Machine	0.536	59.77%
Decision Tree	0.476	56.67%
Biased Confidence	0.687	63.56%

Biased Confidence Classification Algorithm has shown the best results for this dataset. The cross validation accuracy is highest here among all the classifiers. So for the application processes, which require biased attributes in decision making, this algorithm can show better results than other classifiers.

VI. Discussions and Conclusion

Biased Confidence Classification Algorithm though so far used only in a specific domain, can be modified to do prediction in many other domains. For this, all we have to do is modify the biased weight factor. The parameters used in calculating the weight factor are specific to the domain to which they are applied and so to use BCA in some other domain we just have to include the new domain specific parameters instead of the old ones. For example, we can use this algorithm to predict company's stock rise in a particular quarter. For this, we can have weight parameters as Economy factor (a parameter to indicate current position of economy in general) and Last Year Result (giving emphasis to closest performance of the company). In the applications where it performs the classification, although the learning process is supervised, we get the unsupervised learning [20] results as most important attributes, trivial attributes and the attributes responsible for the particular classification. Knowledge of such factors can play important role in future decision making.

The algorithm is yet to be tested for a wide range of data-sets. We believe the Biased Confidence Classification Algorithm can prove to be efficient for different allotment processes. It still has a wide scope for tuning and modification. To make a mark among the established classifier's the algorithm has to test large data-set problems efficiently.

REFERENCES

[1] Chih-Wei Hsu and Chih-Jen Lin, "A Comparison of Methods for Multiclass Support Vector Machines"IEEE transactions on neural network, Vol 13, NO. 2, March 2002.

[2] C. Cortes and V. Vapnik, "Support-vector network," Machine Learning, vol. 20, pp. 273–297, 1995..

[3] Anne Denton and William Perrizo,"A Kernel-Based Semi-Naive Bayesian Classifier Using P-Trees" Department of Computer Science,North Dakota State University Fargo, ND 58105-5164, USA {anne.denton, william.perrizo}@ndsu.nodak.

[4] http://crsouza.blogspot.com/2010/03/kernel-functions-for-machine-learning.html.

[5] Genton, Marc G. "Classes of Kernels for Machine Learning: A Statistics Perspective." Journal of Machine Learning Research 2 (2001) 299-312./

[6] Hofmann, T., B. Schölkopf, and A. J. Smola. "Kernel methods in machine learning." Ann. Statist. Volume 36, Number 3 (2008), 1171-1220.

[7] http://en.wikipedia.org/wiki/Cross-alidation_(statistics)#cite_note-McLachlan-4.

[8] GJ, McLachlan; K.A. Do, C. Ambroise (2004). Analyzing microarray gene expression data. Wiley.

[9] *Houtao Deng, George Runger, and Eugene Tuv, "Bias of Importance Measures for Multi-valued Attributes and Solutions" Proceedings of the 21st International Conference on Artificial Neural Networks (ICANN 2011)*

[10] http://www.mathworks.in/help/toolbox/stats/br039qw-1.html

[11] Murthy,(1998),Automatic Construction of Decision Trees from Data: A Multi-Disciplinary Survey ,Data Mining and Knowledge Discovery 2:p-345-389

[12] M. Aizerman, E. Braverman, and L. Rozonoer, "Theoretical foundations of the potential function method in pattern recognition learning", Automat. Remote Contr., vol. 25, pp.821 - 837 , 1964.

[13] Scholkopf, B., Kah-Kay Sung, Burges, C.J.C., Girosi, F.,Niyogi, P., Poggio, T., and Vapnik, V. "Comparing support vector machines with Gaussian kernels to radial basis function classifiers"Signal Processing, IEEE Transactions on Vol.45 Issue:11 p-2758-2765

[14] http://homepage.cs.uri.edu/faculty/hamel/courses/2011/spring2011/csc581/lecture-notes/15-kernels.pdf

[15] Zhang S.,ZhangC.,Yang Q.(2002),"Data Preparation for Data Mining, Applied Artificial Intelligence,Vol. 17,pp. 375-381

[16] An A.,Cercone N.,(2000),"Rule Quality Measures Improve the Accuracy of Rule Induction" Vol. 1932,pp 119-129

[17] Jiawei Han and MichelineKamber, 2nd Edition. (2006) 'Data Mining:Concepts and Techniques', Chapter 6- Classification and Prediction,pp.285-351

[18] JinhyoKim,"Iterated Grid Search Algorithm on UnimodalCriteria"June, 1997 Blacksburg, Virginia

[19] Abdul RazakHamdan, HamidahJantan and Zulaiha Ali Othman,(2008) 'Data Mining Classification Techniques for Human Talent Forecasting' , pp 2-10

[20] Jennifer G. Dy and Carla E. Brodley. "Feature Selection for Unsupervised Learning" Pub.The Journal of Machine Learning Research archive,Vol. 5, 12/1/2004