Intelligent Rule Mining Algorithm for Classification over Imbalanced Data

Veenu Mangat
University Institute of Engineering and Technology,Panjab University, Chandigarh, India
E-mail:veenumangat@yahoo.com

Dr. Renu Vig
University Institute of Engineering and Technology,Panjab University, Chandigarh, India
E-mail: renuvig@hotmail.com

Abstract—**Association rule mining for classification is a data mining technique for finding informative patterns from large datasets. Output is in the form of if-then rules containing attribute value combinations in antecedent and class label in the consequent. This method is popular for classification as rules are simple to understand and allow users to look into the factors leading to a specific class label. Rule mining methods based on swarm intelligence, specifically particle swarms, can effectively handle problems with large number of instances and mixed data. But the issue of classification over imbalanced datasets, wherein samples from one class greatly outnumber the other class, has not been fully investigated so far. A rule mining method based on Dynamic Particle Swarm and Ant Colony Optimizer that can handle data imbalance, has been proposed in this paper. Performance of the proposed algorithm has been compared with other state-of-the-art methods. Results indicate that in terms of quality, the proposed method outperforms other state-of-the-art methods.**

Index Terms—**association rule mining; classification; PSO; Imbalanced Dataset**

I. Introduction

Data mining refers to the technique of analyzing large data sets in search for interesting and informative patterns. Association Rule Mining is a specific data mining functionality that uncovers interesting patterns or association rules from datasets. An association rule is an implication of the form A=>B, where the L.H.S. A is called the antecedent and R.H.S. B is called the consequent. There is always a measure of certainty or probability associated with the implication. A and B can denote individual itemsets i.e. term containing attribute-value combinations, or conjunction of multiple terms. If the consequent of rule is a class label, the association rule thus mined, can be used to perform classification. This process is called associative classification.

The first application of rule mining for classification was CBA [1]. Since then many other rule based classifiers have been proposed in literature [2-9]. Most of them discover rules using the support-confidence framework. Support indicates the ratio of number of records containing A and B to total number of records in the database. Confidence indicates the probability of B conditioned on A or ratio of number of records containing A and B to number of records containing only A. So support indicates coverage and confidence indicates certainty. The rule based classifier is composed of a collection of rules ranked using confidence or a similar metric.

In real life datasets, often the data is highly imbalanced and skewed, and the instances of one class are far more in number than the other classes. Additionally, the smaller class is the class of interest. Examples of such applications are medical diagnosis [10], risk management [11], credit card fraud detection [12], detection of spillage of oil using satellite imagery [13], etc. Unfortunately, the conventional classification and rule mining methods do not perform well in these cases.

Soft computing is a paradigm that uses inexact solutions to solve problems that are intractable to solve using conventional methods. It is tolerant of imprecision, uncertainty, partial truth, and approximation. It finds an approximate solution that is low cost and guaranteed to be found, in situations where the exact optimal solution is expensive or impossible to find. Evolutionary methods of soft computing paradigm perform well as they do not make any assumptions about the underlying problem. They can deal with vast search spaces and produce near optimal solutions in diverse fields like engineering, biology, genetics, economics, etc. An Evolutionary Algorithm (EA) is a generic population-based meta-heuristic optimization algorithm. It uses some concepts derived from evolution of biological beings like reproduction, mutation, recombination and selection. Swarm Intelligence refers to the collective behavior of multiple agents that act in a decentralized but self organized system in order to find a solution to optimization problems. This behavior is inspired from nature especially behavior of various biological species. Swarm intelligence has been used in various applications like planetary mapping, nanobots, telecommunication,

doi:10.4304/jetwi.6.3.373-379

locating tumors, etc. Specifically, the particle swarm optimizer has given promising results compared to other methods [14]. These techniques can be used to design a rule learning classifier to find a target solution to the problem of class imbalance.

II. Conventional Rule Mining Techniques

Imbalance of data in real life datasets was identified as an important issue for rule mining and classification systems in [15][16]. It has been reported in [17] that there is a bias in learning towards the majority class (class whose samples in dataset far outnumber the other class; other class is called minority class). Classifiers obtain higher predictive accuracy over the majority class, but very low predictive accuracy over the minority class. At times, the predictive accuracy over the minority class is zero because the samples are treated as noise by the learning algorithm. Some classification algorithms fail to deal with imbalanced datasets completely [18][19] and classify all test samples as belonging to majority class irrespective of the feature vector.

To overcome this problem, some algorithms eliminate samples of majority class or modify the class distribution by generating artificial data of minority class. These are called the external approaches. The former approach involves removing samples of majority class, also known as undersampling. It requires identification of non informative samples as a first step. This can be done using various methods like k-nearest neighbor [20] or some evolutionary method [21].

An example of the latter approach is SMOTE (Synthetic Minority Over Sampling Technique). SMOTE generates artificial records lying in between two records of minority class [22]. Borderline SMOTE is a modified SMOTE in which the artificial data lies only on points of the hyperplane separating majority and minority class [23]. Internal approaches are algorithm specific and modify the algorithm to handle class imbalance. External approaches are more general and preferred.

In SMOTE, the minority class is over-sampled by taking each minority class sample and introducing artificial samples along the line joining any of the k minority class nearest neighbors. The number of neighbours that are randomly chosen from the k-nearest neighbours, depends on amount of oversampling required. So the artificial data is created by the randomized interpolation. In order to select a random point along the line segment between two specific samples, artificial samples are generated in the following way:

a) Compute difference between given sample (its feature vector) of minority class and its nearest neighbour.

b) Generate a random number between 0 and 1, and multiply this number by the difference.

c) Add this product to the feature vector of sample under consideration.

Thus, new minority class examples are formed by interpolating between several minority class examples that lie together.

An internal technique of handling skewed data is based on modifying training phase, and it includes ensemble based approaches which make use of under- and oversampling methods to construct class-unbiased base classifiers. A method that combines the benefits of boosting with multiple sampling procedure using SMOTE is known as SMOTEBoost [24]. Another technique is to include artificial data sampling in the process of constructing bagging-based ensemble [25]. Some of ensemble solutions make use of undersampling techniques to construct multiple balanced base learners [26][27].

Imbalanced datasets have also been dealt with by using active learning [28]. Application of active learning techniques for imbalanced data is based on the assumption that the data points accumulated near the borderline are distributed in a more equalized manner than the points in the entire dataset.

Assigning different weights to examples in dataset to reflect their significance, is another method of handling imbalance. Few minority examples can be assigned larger weights to increase their significance compared to majority class examples. Many cost-sensitive methods make use of ensemble classifiers which update the weights while constructing base learners. The weights of minority examples are updated to a greater extent than examples from majority class if they are misclassified by base classifiers. Some of these approaches are: CSB2 [29], RareBoost [30], AdaC1, AdaC2, AdaC3 [31].

Cost-sensitive techniques have been used with Support Vector Machines (SVM) wherein misclassification costs for classes are considered in learning criterion. Cost values can be included in penalization term of learning criterion to construct balanced SVM [32] or SVM with boosting [33].

Several classification approaches like SVM, neural networks, etc., achieve high predictive accuracy. But for the system to be widely adopted, it should give output that is understandable and easily interpretable. Eg: rule based systems. There is considerable overhead involved in data preprocessing and oversampling and undersampling may not lead to a true representative sample in the dataset. Therefore, there is a need to design an algorithm for classification using association rules that can exhibit good performance over balanced and imbalanced datasets without the need for a sampling phase.

Various rule based classifiers using swarm intelligence inspired techniques have been proposed in literature and applied to real world problems in different domains [34-41]. Some of the applications are: advanced swarm intelligence mining algorithm for selecting candidates for surgery for temporal lobe epilepsy [42], association rule mining for discovering hyperlipidemia form biochemistry blood parameters [43], combined PSO and ACO approach to mine data for use in a pharmacovigilance context [44], classification of images [45], flow shop scheduling using discrete Artificial Bee Colony and hybrid differential evolution algorithm [46], etc.

In this paper, we propose an algorithm using hybrid of Particle Swarm Optimization (PSO) and Ant Colony Optimization (ACO) that can handle class imbalance by virtue of its learning algorithm and flexible parameters. So it is an internal approach. Further, performance of proposed algorithm has been compared with other state-of-the-art algorithms for handling imbalanced datasets.

III. PROPOSED METHOD FOR ASSOCIATION RULE MINING

Hybrid algorithms which combine concepts from ACO and PSO can deal with all types of attributes. These methods have reasonably good accuracies while maintaining the comprehensibility of the rules as measured using size of rule and rule sets. To tailor the system to handle dataset imbalance, a new fitness function can be embedded into the process.
The proposed algorithm: Hybrid Optimization based on Swarms (HyS) is based on ACO/PSO and uses a sequential covering approach to discover classification rules one by one.

Pseudocode of HyS:

Algo Main
BestRS = {}
CurrentBestRS = {}
for (i=1; i <=m; i++)
{
TDB= {Training dataset}
If (Remaining training records of class C_i > (MaxUncovExampPerClass * $|C_i|$))
{
Execute **CategAttr**
Execute **ModifiedPSO**
Clip FitRule
CurrentBestRS = CurrentBestRS ∪ FitRule
TDB = TDB −{training records covered by FitRule}
}
}
Clip CurrentBestRS
Sort rules in CurrentBestRS by descending Quality
If Quality(CurrentBestRS)>Quality(BestRS)
{
BestRS= CurrentBestRS
}
Return BestRS

Categorical attributes are handled by the Algo **CategAttr:**

Initialise individuals in population
k=1
While (k< MaxInterations)
{
For every particle *i*
{
Set Rule R_i = "If {null} THEN C"
For every attribute *a* of *i*
{
Apply roulette selection and set state to 0 or1
If (state=1) then
add respective attribute-value pair to R_i;
}
Calculate Quality Q_i of R_i
P = *i*'s past personal best state
Q_p = P's quality
If Q_i>Q_p
{Q_p = Q_i
P = *i*}
}
For every particle *i*
{
P = *i*'s past personal best state
N = best state of any neighbor of *i*
For every attribute *a* of *i*
{
If $P_a = N_a$
Increase pheromone of N_a in the current i_a by Q_p
Else if P_a = off AND seeding term for $i_a \neq N_a$
Increase pheromone for state=0 in i_a by Q_p
Else
Increase pheromone in current i_a of N_a by Q_p
}
Normalize pheromone entries
}
k=k+1
Return fittest rule discovered

Algo ModifiedPSO:
{
Initialise c1=1.4995, c2=1.4995, ω=0.5, V_{max}, V_{min} ;
//initialize parameters
For (n=1; n<=N; n++) //Total particles =N
{
If (count>p)
//refresh period for learning vector formation
{
while (a<D)
{
Generate random number R;
If (R<P_c)
//random number is less than learning probability
{
Select random j □ {i+1, i-1};
Set LVN^a_i =j;
F1=F(LVN^a_i) ;
//Compute fitness of chosen particle
If (F1<F(i^a)) LV^a_i= LVN^a_i
//Particle learns from itself
Else LV^a_i = i;
//assign particle with better fitness to learning vector of current i^{th} particle
}
a=a+1; //Goto next attribute
}
count=count+1;
}
For (a=1; a<=D; a++) //run loop for all attributes
{
$v_i^a = \chi\,(v_i^a + c1\varphi1(P_{ia}^{\,fia} - x_i^a) + c2\varphi2(P_g^{\,a} - x_i^a))$
//Update vel. of i^{th} particle in the a^{th} dim.
$x_i^a = x_i^a + v_i^a$;
//Update pos. of i^{th} particle in the a^{th} dim.

```
        }
    If (F(x_i) < F(pbest_i)
        pbest_i = x_i;
    If (F(x_i) < F(gbest))
        gbest=x_i;
   }                          //Run loop for all particles
if (mod(j,r)==0) regroup; //r is regroup period
}
```

Execution begins with Algo Main that uses a separate-and-conquer method called sequential covering, to mine rules for a specific class at at time. The categorical or nominal attributes are handled by Algo CategAttr using Ant Colony Optimization. The rule discovery for categorical attributes and clipping of individual terms is the same as for ACO/PSO with PF [14]. Continuous attributes are handled by Algo ModifiedPSO.

In AlgoMain, the parameter MaxUncovExampPerClass specifies maximum number of examples of class under consideration, that can be ignored during rule discovery. It is expressed as a percentage of the total number of examples of that class. This parameter is the feature that helps to deal with class imbalance as its value can be adjusted to ensure that some rules cover minority class examples too.

In Algo ModifiedPSO that handles continuous attributes, a local neighbourhood based learning method has been combined with global best learning. The exemplar particles are chosen from a prespecified region rather than randomly. The velocity of a particle is updated according to a vector from its own region only. But the regions are formed again at fixed prespecified points of time in the execution of the algorithm. This mechanism ensures a good balance between the exploration and exploitation properties of the algorithm and avoids premature or delayed convergence.

We construct a vector for each particle which indicates which other particle's personal best should this particle learn from. $LV_i = |LV_i^1, LV_i^2 \ldots LV_i^a|$.

If the fitness of a particle does not increase for a fixed number of iterations (parameter *iter*) then a random number between 0 and 1 is chosen. If this number is greater than P_c, then $LV_i^a = i$. If it is less than P_c, then the particle *i* learns from some other particle's personal best in the same region as given by vector LV_i. P_c is parameter that controls how frequently learning occurs.

To define regions, Von Neumann topology has been considered in this paper with region length as 4. So each particle has two neighbours. The vector to be optimized consists of two terms or dimensions for every continuous attribute that specifies the range for this attribute. Everytime the fitness evaluation of particle is done, the vector is transformed to a set of terms that are *added to Rule produced* by the algorithm. Updation of particle's position and velocity is done for those dimensions using (1)(2):

$$v_i^a = \chi\,(v_i^a + c1\varphi1(P_{ia}^{fia} - x_i^a) + c2\varphi2(P_g^a - x_i^a)) \quad (1)$$

$$x_i^a = x_i^a + v_i^a \quad (2)$$

where, v_i^a is the dimension *a* velocity, x_i^a is the particle position, P_{ia}^{fia} denotes the corresponding dimension *a* of the *i*th particle's own pbest or the exemplar's pbest, P_{ga} is the best position in the neighborhood, χ is constriction coefficient, $\varphi1$ and $\varphi2$ are random weights, c1 and c2 are constants.

A particle operates within its own region. A random particle is picked as seed initially. Other particles set their initial values to a uniformly distributed position between the value of this former seed's continuous attribute and add it to the range for that attribute (for upper bound) and at a uniformly distributed position between seed's value and deduct it from range for that attribute (for lower bound).

Quality, Q of a rule or in other words, fitness of particle is computed using sensitivity and specificity (3):

Specificity X Sensitivity i.e. (TP/TP+FN) (TN/TN+FP)(3)

TP: True positives i.e. number of records covered by the rule that actually belong to the class predicted by the rule.

FP: False positives i.e. number of records covered by the rule that do not belong to the class predicted by the rule.

TN: True negatives i.e. number of records not covered by the rule that do not belong to the class predicted by the rule.

FN: False negatives i.e. number of records not covered by the rule that belong to the class predicted by the rule.

S: Sum of all i.e. (TP + FP + TN + FN) or in other words, the total number of examples.

Since this quality evaluation considers both positives and negatives of both majority and minority class, it is expected to perform well for imbalanced data too. It maximizes the accuracy of each of the two classes with a good balance.

The second modification is to compute quality of not just the individual rules in rule set, but also compute and evaluate the complete rule set quality. This is done using predictive accuracy measure:

Predictive Accuracy= (TP+TN)/S (4)

Predictive accuracy gives the ratio between correct classsifications to the total number of records.

IV. Experimental Setup

A. Dataset Details

We have used eight publicly available datasets to check the performance of our algorithm [47]. These datasets contain a good mix of binary, nominal and continuous attributes. The first four datasets have low clas imbalance and the next four datasets are highly imbalanced. The characteristics of the datasets are as shown in Table I below.

Table I.
Dataset Information

S. No.	Dataset	No. of Attributes	No. of Examples	% age Class Distribution
1.	E-coli0vs1	7	220	35, 65
2.	Wisconsin	9	683	35, 65
3.	Pima	8	768	34.84, 66.16
4.	Hypothyroid2	5	215	16.89, 83.11
5.	E-coli4	7	336	6.74, 93.26
6.	Shuttle2vs4	9	129	4.65, 95.35
7.	Yeast4	8	1484	3.43, 96.57
8.	Abalone19	8	4174	0.77, 99.23

B. Parameters and Results Evaluation

The parameter chosen for evaluation of performance of algorithms is Geometric Mean (GM) of the true rates given as:

$$GM= \sqrt{(TP/TP+FN)(TN/FP+TN)} \quad (5)$$

The accuracy measure does not consider the number of correct labels of different classes, and may be very high even if no rule covers minority class. Hence GM is chosen as it gives weightage to both positive and negative classes.

Similar parameter settings have been used as in hybrid ACO/PSO with PF [3]. For the ACO component, parameter values were set as given: Number of Ants = 500, maximum uncovered examples of each class = 0.1% and number of rules to test if ant has converged = 20. For PSO component, number of particles = 30 and number of iterations = 50. Typical values were set for constriction factor χ = 0.729, social and personal learning coefficients, c1 = c2 = 1.4995. We assume a Von Neumann topology where the length of region is taken as 4. Regrouping is done after every 7 iterations. The learning probability P_c is varied from 0 to 0.5. *Iter* parameter is set to 15.

Comparison has been done with the following state-of-the-art rule mining/classification algorithms: Fuzzy E-algorithm [48], multiobjective fuzzy genetic algorithm (MOFG) [49], C4.5 decision tree algorithm [50], Hierarchical Fuzzy Rule Based Classification System (HFRBCS) [51], BoostedSVM for imbalanced data (BSI) [52]. A standard ten fold cross validation approach has been used. Each dataset is partitioned into 10 subsets, 90% for training and 10% for testing in a round robin fashion for 10 runs. For each data-set, the average result of the ten partitions and runs is reported. The average GM over 10 runs is shown in Table II for each method over each dataset. Values in bold indicate the best value achieved. Table II indicates the results for Geometric Mean of true rates over these eight datasets for these six algorithms.

TABLE II
AVERAGE GEOMETRIC MEAN OF TRUE RATES

Algorithm / Dataset	E-Algorithm	MOFG	C4.5	HFRBCS	BSI	Proposed Hys
E-coli0vs1	95.5	96.7	97.95	93.7	98.3	**99.2**
Wisconsin	96.1	95.8	95.44	88.24	97.3	**98.6**
Pima	55.1	71.11	71.3	68.72	74.67	**75.3**
Hypothyroid2	88.57	94.3	96.5	**99.6**	98	96.9
E-coli4	92.5	86.92	81.38	93.1	92.6	**93.5**
Shuttle2vs4	**100**	99.17	99.2	97.48	91.3	99.2
Yeast4	32.16	71.36	65	82.78	81.4	**83.21**
Abalone19	0	66.09	15.50	70.1	**76.6**	68.11

Results indicate that the proposed Hys gives the highest geometric mean value over five of the eight datasets. This high value is desirable since it implies that the proposed algorithm can mine rules that have good predictive power over both majority and minority class. It consistently shows good performance over all datasets irrespective of the degree of imbalance and is robust. The reason for this is the dynamic learning technique of Hys wherein a good balance between exploration and exploitation of attribute search space is maintained as well as the nature of quality evaluation functions and parameter to control coverage of examples. Thus the proposed technique has the potential to give promising results over both balanced as well as imbalanced datasets.

V. CONCLUSION

This paper discusses a new algorithm that discovers association rules for classification using swarm intelligence. The main motivation is to design a technique that can exhibit good performance over datasets with imbalanced or skewed data distribution of classes. In the proposed Hybrid Optimization based on Swarms (HyS) algorithm, a combination of ACO and PSO techniques is used to mine association rules. ACO handles the nominal attributes and PSO handles continuous attributes. PSO has been modified to learn from both local neighbourhood as well as globally, and to do so dynamically by reconstructing regions. The quality of the not just the rules, but the entire rule set is evaluated according to criteria tailored for this problem. A new parameter ensures that examples of the minority class are also covered by the rule discovery procedure. Results indicate that proposed method performs better or comparably in terms of geometric mean parameter than the other state-of-the-art methods, over both slightly as well as highly imbalanced datasets.

In the future, work can be done on tuning the hyper parameters of the swarm and executing it over imbalanced multiclass datasets. The performance of proposed algorithm can be compared with other associative classifiers using size of the rules and rule sets as parameters. Different quality functions can be used and their effect on geometric mean and rule size can be investigated.

REFERENCES

[1] B. Liu, W. Hsu, and Y. Ma, “Integrating classification and association rule mining”, *Knowledge Discovery and Data Mining*, pp. 80–86, 1998.

[2] W. Li, J. Han, and J. Pei, “CMAR: Accurate and efficient classification based on multiple class-association rules”, in Proc. of the 2001 IEEE *International Conference on Data Mining*, pp. 369–376, Washington, DC, USA, 2001, IEEE Computer Society.

[3] M.-L. Antonie and O. R. Zaiane, “An associative classifier based on positive and negative rules”, in Proc. of the 9th *ACM SIGMOD workshop on Research Issues in Data Mining and Knowledge Discovery(DMKD-04)*, pp. 64–69, 2004.

[4] X. Yin and J. Han, “CPAR: Classification based on predictive association rules”, in Proc. of the *SIAM International Conference on Data Mining(SDM ’03),* D. Barbara and C. Kamath, editors, 2003.

[5] J. Wang and G. Karypis, “Harmony: Efficiently mining the best rules for classification”, in Proc. of the *SIAM International Conference on Data Mining(SDM’05)*, 2005.

[6] G. Cong, A. K.H.Tung, X. Xu, F. Pan, and J. Yang, “Farmer: Finding interesting rule groups in microarray datasets”, in Proc. of the *23rd ACM SIGMOD*

International Conference on Management of Data, pp. 145–154, 2004.

[7] G. Cong, K.-L. Tan, A. K.H.Tung, and X. Xu, "Mining top k covering rule groups for gene expression data", in Proc. of the *ACM SIGMOD/PODS 2005,* pp. 670–681, 2005.

[8] B. Arunasalam and S. Chawla, "Cccs: a top-down associative classifier for imbalanced class distribution", in Proc. of the *12th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 517–522, New York, USA, 2006. ACM Press.

[9] A. Veloso, W. M. Jr., and M. J. Zaki, "Lazy associative classification", in Proc. of the *IEEE ICDM*, pp. 645–654, Los Alamitos, CA, USA, 2006. IEEE Computer Society.

[10] P. Campadelli, E. Casiraghi, G. Valentini, "Support vector machines for candidate nodules classification", *Letters on Neurocomputing*, vol. 68 (2005), pp. 281–288.

[11] Y.M. Huang, C.M. Hung, H.C. Jiau, "Evaluation of neural networks and data mining methods on a credit assessment task for class imbalance problem", *Nonlinear Analysis: Real World Applications*, vol. 7 (4) (2006), pp. 720–747.

[12] T. Fawcett, F.J. Provost, "Adaptive fraud detection", *Data Mining and Knowledge Discovery*, vol. 1 (3) (1997), pp. 291–316.

[13] M. Kubat, R.C. Holte, S. Matwin, "Machine learning for the detection of oil spills in satellite radar images", *Machine Learning*, vol. 30 (2–3) (1998), pp. 195–215.

[14] V. Mangat, "Natural intelligence based knowledge discovery for medical practitioners", *Int. Automation Syst. Eng. Lecture Notes Electrical Eng.*, 2011, vol. 103, pp. 53-64.

[15] N.V. Chawla, N. Japkowicz, A. Kolcz, Editorial: special issue on learning from imbalanced data-sets, *SIGKDD Explorations*, vol. 6 (1) (2004), pp. 1–6.

[16] Q. Yang, X. Wu, "10 challenging problems in data mining research", *International Journal of Information Technology and Decision Making*, vol. 5 (4) (2006), pp. 597–604.

[17] G.M. Weiss, "Mining with rarity: a unifying framework", *SIGKDD Explorations*, vol. 6 (1) (2004), pp. 7–19.

[18] N. Japkowicz, S. Stephen, "The class imbalance problem: a systematic study", *Intelligent Data Analysis*, vol. 6 (5) (2002), pp. 429–450.

[19] C. Phua, D. Alahakoon, V. Lee, "Minority report in fraud detection: classification of skewed data", *SIGKDD Explorations Newsletter,* vol. 6 (1) (2004), pp. 50–59.

[20] I.Zhang, J.Mani, "KNN approach to unbalanced data distributions : a case study involving information extraction", in Proc. of *International Conference on Machine Learning (ICML 2003)*, Workshop Learning from Imbalanced Data Sets, 2003.

[21] S. García, A. Fernández, F. Herrera, "Enhancing the effectiveness and interpretability of decision tree and rule induction classifiers with evolutionary training set selection over imbalanced problems", *Applied Soft Computing*, vol. 9 (2009), pp.1304–1314.

[22] N.V. Chawla, K.W. Bowyer, L.O. Hall, "SMOTE: Synthetic Minority Over-sampling Technique", *Journal of Artificial Intelligence Research*, vol. 16 (2002), pp. 321–357.

[23] H. Han, W.-Y. Wang, B.-H. Mao, "Borderline-SMOTE: a new over-sampling method in imbalanced data sets learning" in: D.-S. Huang, X.-P. Zhang, G.-B. Huang (Eds.), *Advances in Intelligent Computing*, 2005, pp.878–887.

[24] N. Chawla, A. Lazarevic, L. Hall, K. Bowyer, "Smoteboost: improving prediction of the minority class in boosting", In: N. Lavrac, D. Gamberger, H. Blockeel, L.Todorovski (Eds.), *Knowledge Discovery in Databases: PKDD 2003*, Springer, Berlin Heidelberg, 2003, pp. 107–119.

[25] S. Wang, X. Yao, "Diversity analysis on imbalanced data sets by using ensemble models", in Proc. of *2009 IEEE Symposium on Computational Intelligence and Data Mining*, 2009, pp. 324–331.

[26] E. Chang, B. Li, G. Wu, K. Goh, "Statistical learning for effective visual information retrieval", in Proc. of *IEEE International Conference on Image Processing,* vol. 3, 2003, pp. 609–613.

[27] D. Tao, X. Tang, X. Li, X. Wu, "Asymmetric bagging and random subspace for support vector machines-based relevance feedback in image retrieval", *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 28 (2006), pp.1088–1099.

[28] S. Ertekin, J. Huang, L. Bottou, L. Giles, "Learning on the border: active learning in imbalanced data classification", in Proc. of the *30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, ACM, New York, 2007, pp. 127–136.

[29] K. Ting, "A comparative study of cost-sensitive boosting algorithms", in Proc. of the *Seventeenth International Conference on Machine Learning (ICML 2000),* Morgan Kaufmann, 2000, pp. 983–990.

[30] M. Joshi, V. Kumar, R. Agarwal, "Evaluating boosting algorithms to classify rare classes: Comparison and improvements", in Proc. of the 2001 *IEEE International Conference on Data Mining*, IEEE, Los Alamitos, 2001, pp. 257–264.

[31] Y. Sun, M. Kamel, A. Wong, Y. Wang, "Cost-sensitive boosting for classification of imbalanced data", *Pattern Recognition*, vol. 40(2007), pp. 3358–3378.

[32] K. Veropoulos, C. Campbell, N. Cristianini, "Controlling the sensitivity of support vector machines", in Proc. of the *16th International Joint Conference on Artificial Intelligence, (IJCAI999),* Workshop ML3, vol. 1999, 1999, pp. 55–60.

[33] [33] B. Wang, N. Japkowicz, "Boosting support vector machines for imbalanced datasets", *Knowledge and Information Systems*, vol. 25(2010), pp.1–20.

[34] Y. Liu, Z. Qin, et al, "Rule Discovery with Particle Swarm Optimisation", *AWCC 2004, LNCS 3309*, 2004, pp. 291–296.

[35] H. Kwasnicka and K. Switalski, "Discovery of association rules from medical data-classical and evolutionary approaches", in Proc. of the Meeting of *Polish Information Processing Society* Conference, 2005, pp. 163-177.

[36] F. Verhein, S. Chawla, "Using Significant, Positively Associated and Relatively Class Correlated Rules for Associative Classification of Imbalanced Datasets", in Proc. of *Seventh IEEE International Conference on Data Mining*, 2007, pp. 679-684.

[37] R. Cattral, Oppacher F. and K.J.L. Graham, "Techniques for evolutionary rule discovery in data mining", in Proc. of the *IEEE Congress on Evolutionary Computation*, May 18-21, 2009, IEEE Xplore Press, Trondheim, pp. 1737-1744.

[38] X. Yan, C. Zhang and S. Zhangb, "Genetic algorithm-based strategy for identifying association rules without specifying actual minimum support", *J. Expert Syst. Applications*, 2009, vol 36, pp. 3066-3076.

[39] A. Fernández, María José del Jesus, F. Herrera, "Hierarchical fuzzy rule based classification systems with genetic rule selection for imbalanced data-sets", *International Journal of Approximate Reasoning*, 2009, vol 50 , pp. 561–577.

[40] E. Gonzales, K. Taboada, S. Mabu, K. Shimada and K. Hirasawa, "Combination of two evolutionary methods for

mining association rules in large and dense databases", *J. Adv. Comput. Intell. Intelligent Inform.,* 2010, vol. 13, pp. 561-572.

[41] J. Vreeken, M. van Leeuwen and A. Siebes, "KRIMP: Mining itemsets that compress", *Data Mining Knowledge Discovery J.*, 2011, vol. 23, pp. 169-214.

[42] M. Ghannad-Rezaie, H. Soltanain-Zadeh, M.R. Siadat and K.V. Elisevich, "Medical data mining using particle swarm optimization for temporal lobe epilepsy", in Proc. of the *IEEE Congress on Evolutionary Computation,* Jul. 16-21, 2006, IEEE Xplore Press, Vancouver, BC, pp. 761-768.

[43] S. Dogan and I. Turkoglu, "Diagnosing hyperlipidemia using association rules", *Mathem. Computational Applications*, vol. 13, 2008, pp. 193-202.

[44] M. Sordo, N. Shawn and A. Murphy, "PSO/ACO approach to knowledge discovery in a pharmacovigilance context", in Proc. of the *11th Annual Conference Companion on Genetic and Evolutionary Computation* Conference: Late Breaking Papers, (LBP '09), ACM New York, NY, USA, pp. 2679-2684.

[45] N. Wahid, "A novel image classification algorithm using swarm-based technique for image database", *Ubiquitous Comput. Multimedia Applications,* Springer-Verlag J., 2011, vol. 151, pp. 460-470.

[46] M. F. Tasgetiren, Quan-Ke Pan, et al, "A discrete artificial bee colony algorithm for the total flowtime minimization in permutation flow shops", *Information Sciences*, vol 181, 2011, pp. 3459–3475.

[47] UCI machine learning repository. University of California, Irvine, School of Information and Computer Sciences. URL:<http://www.ics.uci.edu/~mlearn/MLRepository.html >.

[48] L. Xu, M.Y. Chow, L.S. Taylor, "Power distribution fault cause identification with imbalanced data using the data mining-based fuzzy classification e-algorithm", *IEEE Transactions on Power Systems*, vol. 22 (1) (2007), pp. 164–171.

[49] H. Ishibuchi, Y. Nojima, "Analysis of interpretability-accuracy tradeoff of fuzzy systems by multiobjective fuzzy genetics-based machine learning", *International Journal of Approximate Reasoning*, vol. 44 (2007), pp. 4–31.

[50] J. Quinlan, C4.5: Programs for Machine Learning. Morgan Kaufmann, 1993, 302 pages.

[51] Alberto Fernández, María José del Jesus, Francisco Herrera, "Hierarchical fuzzy rule based classification systems with genetic rule selection for imbalanced data-sets", *International Journal of Approximate Reasoning,* vol. 50 (2009), pp. 561–577.

[52] M. Zieba, J. M. Tomczak, M. Lubicz, J. Swiatek, "Boosted SVM for extracting rules from imbalanced data in application to prediction of the post-operative life expectancy in the lung cancer patients", *Applied Soft Computing*, 2014, vol. 14, pp. 99-108.

Veenu Mangat, Masters of Engineering in Computer Science, has been working as an assistant professor in information technology for the last 10 years. She is currently pursuing her PhD in the area of data mining. She has eight publications in national and international journals and has presented five papers in international conferences of repute. Her research areas include data mining, optimization techniques and soft computing.

Prof. (Dr.) Renu Vig, PhD, has a teaching/research experience of more than 26 years. Her areas of specialization are artificial intelligence, neural networks and signal processing. She has guided several PhDs successfully in the area of signal processing, cryptography and fuzzy logic based control. She has published more than 54 papers and coauthored two books. She has been investigator and supervisor for research projects related to software protection and modelling of Nano Scale MOSFETs.