IOSR Journal of Engineering (IOSR JEN) **www.iosrjen.org**
ISSN (e): 2250-3021, ISSN (p): 2278-8719
PP 59-65

Classification Model Using Artificial Bee Colony Optimization Technique

Deoshree D. Diwathe
Department of Computer Science & Engineering, Rashtrasant Tukdoji Maharaj Nagpur University, Maharashtra, India

Abstract: *Data mining is a process of extracting and analyzing the large quantities of the dataset and discovering useful knowledge. Discovered Knowledge is useful for creating predictive structure. Classification Technique is useful to analyze and classify the set of data and generate the classified model. In this research, decision tree classifier with the greedy approach is implemented. With the help of greedy approach, the decision for each and every attributes gives perfect classification result But, the Greedy approach is generating a number of IF-THEN rules and proposed system is getting very complex and taking more time for classifying the data. Artificial Bee Colony optimization algorithm is combining with a Decision tree; it is called as Hybrid Algorithm. It is useful for classification task to overcome this problem.UCI Repository benchmark datasets such as Breast Cancer dataset and Hepatitis dataset are used. Hybrid Algorithm is very powerful, effective and easy to understand. This hybrid proposed model gives optimized rules, also minimized the complexity and maximize the accuracy andefficiency.*

Keywords:*Data Mining, Classification technique, Decision Tree Classifier with the greedy approach, Confusion Matrix, Artificial Bee Colony Optimization Algorithm (ABC Optimization Algorithm).*

I. Introduction

Data Mining is a method to extract potential, valid and implicit knowledge from huge volumes of raw data. The extracted information and knowledge is predictive but must be accurate, easy to understand, readable, comprehensible. In this research paper, the unique algorithm and very effective algorithm used in data mining is called Evolutionary Algorithm (EA) such as Artificial Bee Colony Optimization Algorithm. In this research, proved that how Evolutionary Algorithm for classification process is better than Decision Tree algorithm for classifying the set of raw data.

Classification technique is used in machine learning problem. In this technique, the set of data is divided into two parts, first one is training part and the second one is testing part. The training part is more focused because classification is supervised learning so the training is more efficiently worked out. The training data instances are formed to create the classified model and the testing data instances are formed for validating that classified model (Predicting values from unknownclasses).

Decision Tree Classifier is representing the rules, which are used to classify and analyze the datasets. The Greedy approach is top – to - bottom approach and it gives perfect decision. However, it is generating a number of rules that is why the proposed model is getting very complex and the performance of Classification model is taking time during classifying the data. Therefore, the accuracy is going to decrease.

In this proposed system, Artificial Bee Colony Optimization algorithm (ABC) is combining with the decision tree with a greedy approach, which is used to solve the drawback of Decision Tree Classifier with the greedy approach. A Hybrid algorithm is used to optimize rules, time and complexity of proposed model compared to another model. The organization of this research paper is as follows: - Section 2 contains related work, Section 3 contains description about classification algorithm, Section 4 contains Artificial Bee Colony optimization algorithm, Section 5 contains description about hybrid proposed model, Section 6 contains empirical results, Section 6 contains Conclusion of proposed model.

II. Methodology

Related Work

The following information shows different data mining techniques used in the Classification over different datasets. The unique concept is classifying different dataset usingDecision tree classifier with Optimization Algorithm and increase their efficiency and accuracy.

- **C 4.5 Algorithm [6]**: - "Jia Yu, Yun Chen", the author suggested C 4.5 algorithm is applying onto the training data at that time all training cases are organized in the structure of IF-THEN rule and the number list is rapidlyincreases.

- **Support Vector Machine (SVM) classifier [7]: -** "NoureddineGhoggali et.al." suggested SVM classifier is used to classify the limited set of unlabeled data and genetic algorithm is used for the optimization algorithm. SVM classifier is one of the strong classifier but it is not a solution for a tricky and complex problem duringclassification.

- **Artificial neural network (ANN) [8] [11] [14] [15]:-** Authors "Beatriz A. Garro et.al, Biprodip pal et.al. RozaidaGhazali et.al. And Wang Shiqing et.al." described Artificial neural networks (ANN) is designing a mechanism of error- testing that tests different architectures, transfer functions and the selection of a training algorithm that permits to adjust the synaptic weights of the ANN[8]. Fuzzy ruled- based system is combined with the optimization algorithm. There are some drawbacks such as Fitness of a system is decreases because of generating the number of rule lists and Fuzzy logic worked on the labeled dataset only not on the real dataset[11].

ANN must have a large amount of framework or measures because it generates better accuracy and efficiency of classification [14]. The weights of the input vectors using ANN are very high. That is why; the accuracy of proposed model has been decreased [15].

- **Ant Colony Optimization Algorithm [6] [10]: -** The target of the optimization algorithm is, maximized efficiency and optimized complexity using Ant Colony Optimization [6]. Ant Colony Optimization Algorithm is applying to solve the routing problem in easy or simpler way[10].

- **Artificial Bee Colony optimization algorithm [8] [14] [15]: -** Artificial Bee Colony is applying to increase the accuracy and decrease the complexity i.e. number of connection of the ANN is decreases and problem is easily solved [8]. Artificial Bee Colony Optimization algorithm is having a capability of exploitation and exploration, which is used to update the weights, and solving the problem of ANN classifier in the optimal way [14]. This algorithm is used to optimize the architecture of the complex network and increases the accuracy and efficiency[15].

- **Particle swarm optimization algorithm [9] [12]: -** PSO algorithm for population classification in fire evacuation is used to remove those drawbacks and optimized the recall measures and conditions related to population classification [9]. Particle Swarm Optimization Algorithm is used to generate the optimal result. Particle swarm optimization (PSO) algorithm to find out those nearest points from that group [12].

III. Methods

- **Decision Tree Classifier with Greedy Approach**

Decision Tree Classifier is ruled generating classifier in a tree structure [6]. Decision tree structure is indicating root node, leaf node, andbranches.

- **Root: -** The Root node indicating the topmostnode.
- **Internal Nodes: -** Each internal node is indicating condition, which is applied to anattribute.
- **Branch: -** Each branch of the tree is a representing an output of thatcondition.
- **Leaf Node: -** Each and every leaf node is containing classes.

The Decision tree classifier represented by IF-THEN rules formation. This rule dependent classifier with the greedy approach is used to spread information very easily. So that, the classification of each instance is correctly and perfectly classified the data during execution of IF-THEN rules which is denoted as follows:-

IF < Condition > THEN < Class >

From above example, the IF-THEN rule is containing two parts:-

- **Antecedent: -** The IF part is antecedent consist set of conditions. These conditions are connected by AND operator.
- **Consequent: -** The THEN part is consequent and it specifies the class predicted for cases whose attributes satisfy all the terms specified in the ruleantecedent.

Decision Tree Classifier with Greedy Approach are having many advantages like, this algorithm does not require any domain related information or knowledge, easily create the decision tree structure, easily comprehensible, the implementation of classification and learning or testing steps are executed very fastly and easily. It is generating a number of rules during classification task, that's why the complexity of proposed

classification model becomes high and decreases the accuracy of proposed model. In this research, with the help of hybrid algorithm, an optimized solution is created which is used to overcome this problem and enhance theperformance.

- **Artificial Bee Colony Optimization Algorithm**

Artificial Bee Colony (ABC) optimization algorithm is swarmed intelligence meta-heuristic algorithm [8]. Artificial Bee Colony Optimization Algorithms is invented by Karaboga in 2005. Meta-heuristic ABC algorithm is applying for solving a complex problem. It is mostly applicable for investigating the global optimal solution. It is very simple and effective than any other evolutionary algorithms. In Artificial Bee Colony optimization algorithm are consisting three Types of Honey Bees depending on their own task. Those bees are Employed Bee, Onlooker Bee and Scout Bee[14].

Firstly Employed Bee goes outside from a hive for finding out the food sources. Employed bees are going to different directions for searching the nectar amount and sources of pollens. After completion of searching the sufficient nectar amount, those bees go back to home called as a hive and start their dance called as a "Waggle dance" in the hive[15].

With the help of dance, those bees are spreading information related to nectar amount and sources of pollen. Those bees are also spreading information about the location of food sources, how many quantities is available and also gives information about the quality of food sources. Onlooker Bees and Scout Bees are Unemployed Bees. Onlooker Bees are watching employed Bee's dance and get some profitable information related to the nectar amount. Now onlooker bees are becoming employed bees and go outside from hive for determining the Neighbor and profitable food sources and food sources are selected according to the probability value. Scout Bees are used to finding out new food sources randomly around thenest.

- **Proposed Work**

In proposed work, Artificial Bee Colony Optimization Algorithm is combining with Decision tree with greedy approach. Hybrid algorithm is designing as follows:-

- **Hybrid Algorithm:-**

Step 1:- Initialize and evaluate the population of solution.
Step 2:- Create node N.
Step 3:- Ifall instances are in the same class, **then** Return with the same class.
Step 4:- Attribute_Selection method through Greedy Selection process from training dataset.
Step 5:- Calculate Probability of each instances using following formula:-

$$P(i) = \frac{Fit(i)}{\sum_{n=1}^{SN} Fit(n)} \quad (1)$$

Where,
P (i) = probability value for each instances
Fit (*i*) = fitness value of the solution i
SN =Upper bound
n = Lower Bound
Fit (n) = Mean value of fitness value of i

Step 6:- Generate new rules using lower and upper bound.

$$Vij = Xij + \Phi ij\ (Xij - Xkj) \quad (2)$$

With the help of above formula, we have to calculate lower bound and upper bound of each and every attributes from the dataset as follows:-

- Lower Bound= $f - k1((f(max) - f(min))$ (3)
- Upper Bound = $f + k2\ (f(max) - f(min))$ (4)

Step 7:- Add new generation rules into new rule list.

Step 8:- If probability value is in between lower and upper bound of each attributes **then,** for each outcome instances are under splittingcriterion.

Step 9:- Let set of data instances in population satisfying outcome.

Step 10:- If data instances are empty **then,** Attach a leaf labeled with the majority class **Else** attach an attribute returned by Generate hybrid decision tree to node N.

Step 11:- Return N.

- **Description about Hybrid Algorithm:-**

In above hybrid algorithm, implementation is carried out by combining two algorithms i.e. decision tree algorithm and artificial bee colony optimization algorithm. In hybrid algorithm, the step of attribute selection method is carried out with the help of greedy selectionprocess.

Then calculating probability for each instance and checked this probability under the lower bound and upper bound of each attribute. If set of instances satisfying the condition, then the optimal outcome will begenerate.

- **Flowchart for Hybrid Algorithm:-**

Hybrid algorithm for classification task is implementing by following way. The flow of hybrid algorithm is described using following flowchart for classification using decision tree combining with artificial bee colony optimization algorithm as follows:-

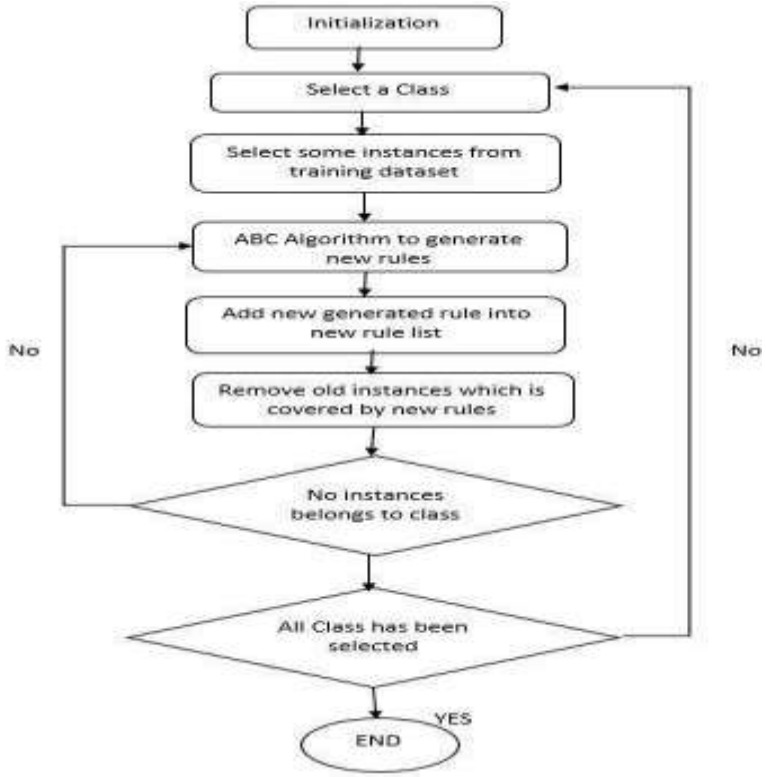

Figure 1: *Flowchart of Hybrid Algorithm*

IV. Performance Evaluation

- **Dataset Used-** In this proposed model, two medical datasets are used for implementation as follows:-

- **Breast-Cancer Dataset:-**
  - Number of Classes:- 2(Benign and Malignant)
  - Number of instances:-699
  - Number of attributes:- 10 + the class attribute
- **Hepatitis Dataset:-**
  - Number of instances:- 155
  - Number of attributes:-20(including class attributes)
  - Number of classes:- 2 (Die and Live)

Confusion Matrix: - Confusion matrix is relational matrix between the Actual class labels and Predicted class labels. Confusion Matrix is used to calculate the correctly classified data and incorrectly classified data.
The main four terms are as follows:-

- **True-Positive (TP):-** We predicted "yes" and Actual value is "yes" means actual number of instances covered by rules are equal to the predicted number of instances covered by rules.
- **False-Positive (FP): -** We predicted "yes" but Actual value is "no", means actual number of instances is not covered by the rule which is rules covered by the predicted class.
- **True-Negative (TN): -** We predicted "no" but actual value is also "no" means actual number of instances is not covered by rule and predicted value of instances not covered by rules.
- **False-Negative (FN):-** We predicted "no" but actual value is "yes" means actual number of instances is covered by rules but predicted value instances is not covered by the rules. Confusion Matrix is designed with above four main terms as follows:-

	Actual Class	
Predicted class	**Class 1**	**Class 2**
Class 1	TP	FP
Class 2	FN	TN

Table1. *Confusion Matrix*

➢ **Results for Breast-Cancer Dataset:**

Confusion matrix for decision tree classifier for Breast Cancer datasets as follows:-

- **Breast Cancer: -** There are two classes such as Benign (non-cancerous) and Malignant (cancerous). There are total 699 instances and 10 attribute and one class attribute.

	Class 1	**Class 2**
Benign	437	76
Malignant	21	165

Table 2. *Confusion Matrix for decision tree classifier with Greedy Approach*

According to above confusion matrix, the correctly, incorrectly classified data instances and Accuracy of decision tree with greedy approach classification algorithm is calculated as follows:-

Dataset	**Correctly Classified Instances**	**Incorrectly classified Instances**	**Accuracy**	**Error Rate**
Breast Cancer	602	97	86.12%	13.87%

Table 3. *Output of Decision Tree with Greedy Approach algorithm*

Confusion matrix for Hybrid algorithm for Breast Cancer datasets as follows:-

	Class 1	**Class 2**
Benign	448	66
Malignant	10	175

Table 4.*Confusion Matrix for Hybrid Algorithm*

According to above confusion matrix, the correctly, incorrectly classified data instances and Accuracy of Hybrid algorithm is calculated as follows:-

Dataset	**Correctly Classified Instances**	**Incorrectly classified Instances**	**Accuracy**	**Error Rate**
Breast Cancer	623	76	89.12%	10.87%

Table 5. *Output of Hybrid Algorithm*

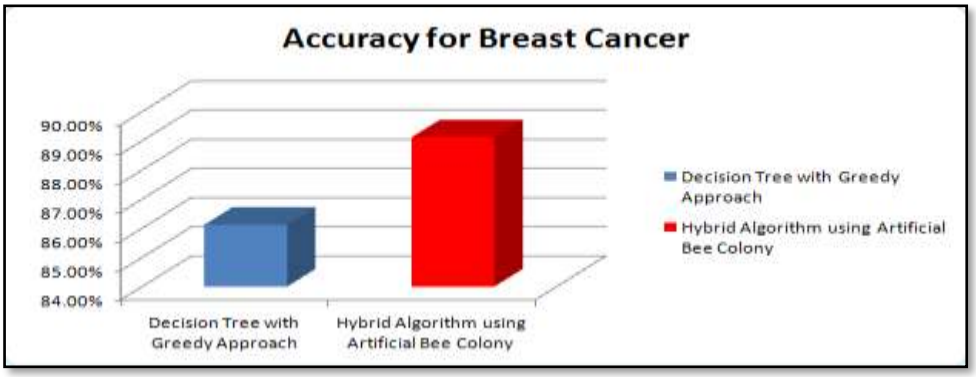

Figure-2.*Comparison of Decision tree with Greedy Approach and Hybrid Algorithm*

➢ **Results for Hepatitis Dataset:-**

Confusion matrix for decision tree classifier with greedy approach for Hepatitis datasets as follows:-

- **Hepatitis Dataset: -** There are two classes such as DIE and LIVE. There are total 155 instances and 20 attribute including one class attribute.

	Class 1	**Class 2**
DIE	0	13
LIVE	32	110

Table 6.*Confusion Matrix of decision tree classifier with Greedy Approach*

According to above confusion matrix, the correctly, incorrectly classified data instances and Accuracy of decision tree classification algorithm is calculated as follows:-

Dataset	**Correctly Classified Instances**	**Incorrectly classified Instances**	**Accuracy**	**Error Rate**
Hepatitis	110	45	70.96%	29.03%

Table 2.*Output of Decision Tree with Greedy Approach*

Confusion matrix for Classification using optimization algorithm for Hepatitis datasets as follows:-

	Class 1	**Class 2**
DIE	03	10
LIVE	29	113

Table 3.*Confusion Matrix for Hybrid Algorithm*

According to above confusion matrix, the correctly, incorrectly classified data instances and Accuracy of classification using optimization algorithm is calculated as follows:-

Dataset	**Correctly Classified Instances**	**Incorrectly classified Instances**	**Accuracy**	**Error Rate**
Hepatitis	**116**	**39**	**74.83%**	**25.16%**

Table 4.*Output of Hybrid Algorithm*

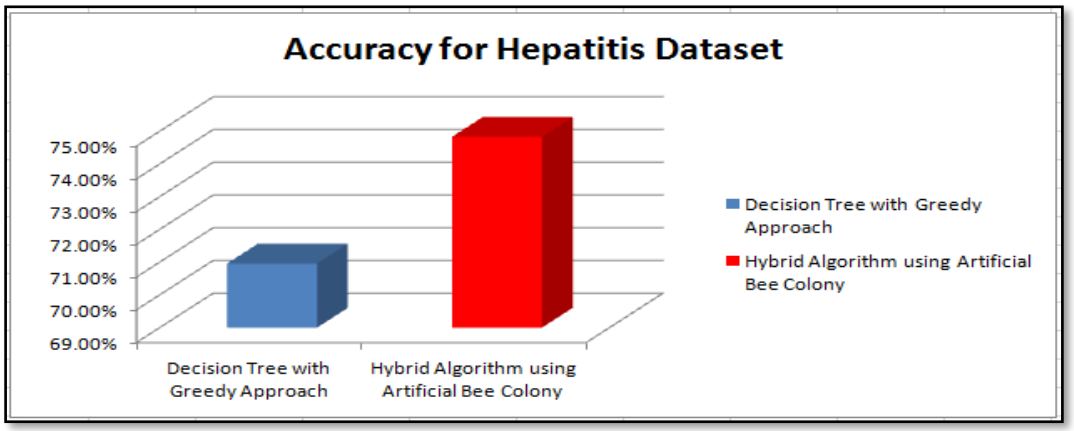

Figure-3. *Comparison of Decision tree with Greedy approach and Hybrid Algorithm*

V. Conclusion

In this research paper, we have concluded that, Artificial Bee Colony is combined with Decision tree classifier and produced hybrid classification model using ABC optimization technique which is well implemented, very efficient and powerful algorithm for classification task. According to Empirical results, Hybrid algorithm is having more accuracy than the decision tree classifier with greedy approach. The proposed Hybrid model reduced complexity of the model, classifying data within short time and maximized the accuracy and efficiency of hybrid model.

References

[1]. Abdul Rauf Baig, Member, IEEE, Waseem Shahzad, and Salabat Khan, "Correlation as a Heuristic for Accurate and Comprehensible Ant Colony Optimization Based Classifiers", IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTATION, VOL. 17, NO. 5, OCTOBER 2013, PP.686-704.

[2]. Ting-Cheng Feng, Tzuu-Hseng S. Li, "Advanced Hierarchical Fuzzy Classification Model Adopting Symbiosis Based DNA-ABC Optimization Algorithm", Scientific Research PublishingInc. March 2016, pg no- 441-455.

[3]. Nadezda STANAREVIC, Milan TUBA, Nebojsa BACANIN, "Enhanced Artificial Bee Colony Algorithm Performance", International Journals of LATEST TRENDS on COMPUTERS (Volume II), page no. 440-445.

[4]. D.Lavanya1 and Dr.K.Usha Rani, "ENSEMBLE DECISION TREE CLASSIFIER FOR BREAST CANCER DATA" , International Journal of Information Technology Convergence and Services (IJITCS) Vol.2, No.1, February 2012,page no.- 17-24.

[5]. Hanning Chen, Lianbo Ma, Maowei He, Xingwei Wang, "Artificial Bee Colony Optimizer Based on -Cycle for Stationary and Dynamic Optimization", IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS: SYSTEMS 2016,pp.1-20.

[6]. Jia Yu, Yun Chen, "Ant Colony Optimization based Land Use Suitability Classification", IEEE International Conference on Intelligent Computing Applications 2016, PP. 1-6.

[7]. NoureddineGhoggali,FaridMelgani,YakoubBazi, "A Multiobjective Genetic SVM Approach for Classification Problems With Limited Training Samples" , IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING JUNE 2009, VOL. 47, NO. 6, PP.1707-1718.

[8]. Beatriz A. Garro , Humberto Sossa ,Roberto A. Vazquez, "Artificial Neural Network Synthesis by means of Artificial Bee Colony (ABC) Algorithm" , IEEE 2011, pp.331-338.

[9]. Yu-Jun Zheng, Hai-Feng Ling, Jin-Yun Xue, Sheng-Yong Chen, "Population Classification in Fire Evacuation: A Multiobjective Particle Swarm Optimization Approach", IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTATION FEBRUARY 2014, VOL. 18, NO. 1, PP. 70-81.

[10]. Rafael S. Parpinelli, Heitor S. Lopes, Alex A. Freitas, "Data Mining With an Ant Colony Optimization Algorithm", *IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTING AUGUST 2002*, VOL. 6, NO. 4, pp. 321-332.

[11]. Biprodip Pal, Mumu Aktar, Firoz Mahmud, Syed TauhidZuhori, "An Evolutionary Fuzzy Genetic and Naïve Bayesian Approach for Multivariate Data Classification", IEEE International Conference on Computer and Information Technology 2014,pp.20-24.

[12]. Alejandro Cervantes, Inés MaríaGalván, and Pedro Isasi, "AMPSO: A New Particle Swarm Method for Nearest Neighborhood Classification", IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART B: CYBERNETICS OCTOBER 2009, VOL. 39, NO. 5, PP.1082-1091.

[13]. M. Duraisamy, F. Mary Magdalene Jane, "CELLULAR NEURAL NETWORK BASED MEDICAL IMAGE SEGMENTATION USING ARTIFICIAL BEE COLONY ALGORITHM", IEEE International Conference on Computer and Information Technology 2012, PP.1-6.

[14]. Yana MazwinMohmadHassim,RozaidaGhazali, "Solving a Classification Task using Functional Link Neural Networks with Modified Artificial Bee Colony", IEEE Ninth International Conference on Natural Computation (ICNC) 2013,pp.189-193.

[15]. LvQiongshuai, Wang Shiqing, "A Hybrid Model Of Neural Network and Classification in Wine" IEEE 2011, PP.58-61.