Cuckoo Optimization and Fuzzy Logic Classifier with an Enhanced Stacking Algorithm

Kalaivani K, Dr.N. Uma Maheshwari
Assistant Professor, Professor, Department of Computer Science & Engineering,
PSNA College of Engineering & Technology,
Dindigul, Tamilnadu.
kkalaivani@psnacet.edu.in.

Abstract — In the medical field accuracy plays an important role as it concerns with the life of an individual. Using machine learning techniques various research has been done to predict and classify diseases. Any how we are not able to detect which classifier produce best result. Also, there is a difference in the ratio of classes due to imbalanced data. We often face these types of data analysis in real life easily. In this research work the imbalanced data classification problem could be addressed in a systematic way by applying enhanced stacking with cuckoo optimization which helps to accelerates the performance of learning algorithms. Also, in this research work we propose a diagnosis system to classify heart disease using cuckoo search optimization to reduce feature with modified stacking also fuzzy logic system is used to predict the heart disease. We prove that ensemble classifier is more accurate than any single component classifiers. Initially the data collected are imbalanced. We balance the data and make it better prediction. To predict Ischemic Heart Disease in more accurate we propose a new ensemble heterogeneous classifier construction process with modified stacking to engender domain-specific configurations. We consider both the field ensemble learning and distributed data mining to create an efficient and enhanced version of the standard stacking ensemble learning techniques by using fuzzy logic and cuckoo optimization creating the meta classifier. The experimental result shows that enhanced stacking with cuckoo optimization outperforms the existing approaches.

Index Terms— Modified Stacking, fuzzy logic system, Cuckoo optimization, Ensemble classifier.

I. INTRODUCTION

Classification is the method of classifying the data. In nature, data is present in varied form thus proper classification of the data is essential to provide better performance for mining valuable information from the data. In machine learning, data classification is the problematic of classifying dataset into many classes or subsets used to achieve several data extraction tasks, for better classification for the new observations a training dataset or a predefined classified dataset is used. Classifiers are nothing but various algorithms are used to classify the data. Depending upon the various parameters and its selection their classification performance is judged. In order to generate better classifiers for the data optimized parameters are mandatory. There are several parameter optimization techniques like particle swarm optimization, Ant colony optimization, and Bee colony optimization etc. are contemporary in nature which used to afford optimized parameter to categorize the data and engender optimized classifiers. But these techniques degrade the performance of the whole system which do not have easy adaptability.

Cuckoo Search algorithm is constructed upon parasitic reproductive approach of some cuckoo species. Cuckoo birds lay their eggs in some other bird’s nest. If those bird becomes aware of these eggs, then sometimes they throw away those eggs or sometimes leave their nest and construct a new one.

Cuckoo Search Pseudocode:
Algorithm 1: Cuckoo Search via Levy Flights
Start
Initialize randomly a population of ‘n’ swarm nests (i.e. solutions) as xi, (i = 1, 2…, n).
Apply problem-specific fitness function, f(x) to evaluate the optimum nest with best fitness value
WHILE (best fitness value < a threshold t1 OR number of iterations < a threshold t2)
Discard a fraction probability [0,1] of the worst nests
Produce new nests by Levy Flight
Substitute discarded nests with new nest using biased random walks
Find current best nest in the current population using f(x)
IF (fitness (current best nest) > fitness (best nest))
Update best nest = current best nest
End IF
End WHILE
Output best nest
End

Cuckoo Search is mostly based on three rules. Firstly, they may choose a random nest and dumps its egg which was laid by each cuckoo. Secondly, their eggs may carry over to the next generation if the nest contains high quality of eggs. At last, there are constant number of host nest and they discover the egg with some probability laid by cuckoo, the aim of cuckoo search is to arrive at optimum solution over several generations.

II. RELATED WORK

[1] In this research work author improve the decision of the classifiers for heart disease diagnosis through an ensemble method. There are two types of ensemble homogeneous and heterogeneous ensemble.in this paper author applied homogeneous ensemble for heart disease classification and outcomes are optimized by using Genetic algorithm. Here by using IO-fold cross validation data is
evaluated and their performance of the system is estimated by classifiers accuracy, sensitivity and specificity to check the possibility of our system. Comparison of our methodology with existing ensemble technique has shown considerable improvements in terms of classification accuracy. For searching quality solution Genetic Algorithm has been found a very good technique for optimization and the proposed framework of SVM classifier ensemble and optimization of results using Genetic Algorithm technique improved the classification accuracy as compared to existing work. [2] In this research work author proposed a genetic algorithm based feature selection for the heart diseases and the details are, to find informative features that play a significant role in discrimination of samples the author presented a genetic algorithm (GA) based feature-selection method. They build a classifier by selecting a subset from multiple Genetic Algorithm. In order to improve the classification performance, he introduced an approach which can be combined with various classifiers to improve and selection of the most discriminative features. Initially with a set of pre-selected features by using a filter, the author used a Genetic Algorithm combined with Fisher’s linear discriminate analysis (LDA) to discover the space of feature subsets. In fact, our proposed method pays GA and uses the LDA classifier to estimate the fitness of a specified candidate feature subset. An external test set was selected by using Kohonen self-organizing maps (SOMs) to estimate the performance of feature selection at the ending stage. The proposed technique can be used to analyze CHD in patients without using any angiographic methods, which may have a high risk of death for the entities. [3] In this research work author discussed on Naive Bayes algorithm method for heart disease, the particulars are data classification is constructed on supervised machine learning algorithms which result in accuracy, this algorithm takes time to build. Author used Tanagra tool to classify the data and also, he evaluated the data using 10-fold cross validation and the results are compared with other algorithms. Based on their performance algorithms are selected, but not around the test dataset itself, and also encompassing the predictions of the classification methods on the test instance. By recording the predictions of each algorithm training data are produced, both for training and for testing using the full training data. Performance is determined by running 10-fold cross-validations and averaging the calculations for each training dataset. Numerous approaches have been planned for the characterization of learning domain. The performance of each algorithm on the data attribute is verified. The algorithms are graded according to their performance of the error rate. He also pacts with the results in the attribute of dataset classification obtained with Naive Bayes algorithm, Decision list algorithm and k-nn algorithm, and on the entire performance made recognized Naive Bayes Algorithm when tested on heart disease datasets. Naive Bayes algorithm is the greatest compact time for processing dataset and demonstrations better performance in accuracy prediction. [4] In this research work author uses optimization techniques such as particle swarm optimization, to proposes a novel system to classify three types of electrocardiogram beats, namely normal beats and two manifestations of heart arrhythmia for a heart disease classification system. In this paper, he divides the system into three main modules namely, a feature extraction module, a classifier module, and an optimization module. In the first module system, feature extraction module, a proper set combining the shape features and timing features is planned as the effective characteristic of the patterns. In the second classifier module, a multi-class support vector machine (SVM)-based classifier is planned. Finally, for the optimization module, a particle swarm optimization algorithm is proposed to examine for the best value of the Support Vector Machine parameters and upstream by looking for the finest subset of features that feed the classifier. The experimental results illustrate that the proposed algorithm has high efficiency recognition accuracy. This high efficiency is realized with only little features, which have been nominated using particle swarm optimizer.

[5] In this research work author proposed a Back propagation neural network and genetic algorithm for heart diseases analysis classification, by using the Three-Term Back propagation (TBP) network created on the Elitist Multiobjective Genetic Algorithm (MOGA). One of the current MOGAs is a Non-Dominated Sorting Genetic Algorithm II (NSGA-II), which is used to reduce or optimize the error rate and network structure of TBP concurrently to attain more accurate classification results. In addition, accuracy, sensitivity, specificity and 10-fold cross validation are used as performance estimation indicators to estimate the result of the proposed method.

III. BACKGROUND

A. Stacking

The Stacking approach (which is the abbreviation of Stacked Generalization) is the approach to combine classifiers from different learning algorithms (Wolpert, 1992). Since different learning algorithms apply different knowledge representations and different learning biases, thus different classifiers will be generated. Consequently, the errors from the classifiers are not closely correlated to each other, and diversity can be expected (Wolpert, 1992).

Fig Standard Stacking Ensemble Learning

Algorithm: Stacking
1. Split the training set S into N partitions, N > 1
2. For each learning algorithm
   (a) Use the leave-one-out training process to train the classifier Ct
   (b) Validate Ct and output the validation result Rt
3. Create the meta training set S0 by joining the Rt and class label: y
4. Train the meta classifier M with S0.

B. Metaheuristic Search Algorithms

Algorithms with stochastic component are often referred to as meta-heuristics. They are also sometimes called as nature-inspired algorithms metaheuristics. Through trial and error, we discover or find something is referred as heuristics. Meta refers to high level or beyond and generally meta-
heuristics do well than simple heuristics. The word “metaheuristics” was coined by Fred Glover in his seminal paper and a meta heuristics can be considered as a “master strategy that guides and modifies other heuristics to produce solutions beyond those that are normally generated in a quest for local optimality” (Glover and Laguna 1997). Most of the metaheuristics algorithms are suitable for global optimization. There are two major components in metaheuristics they are intensification and diversification or exploitation and exploration.

Several metaheuristics algorithms are Particle Swarm Optimization, Firefly, Cuckoo Search, Ant Colony Optimization, Genetic Algorithms, Bee Algorithms, Taboo Search etc. In this research work Cuckoo Search Optimization has been covered.

C. Combination Of Enhanced Stacking Ensemble With Metaheuristic Algorithm

The stacking ensemble learning using Cuckoo algorithm as the meta-learner. The training data set is divided into a training subset and a holdout subset. The training subset is additionally divided into n subsets by means of stratified sampling with replacement, which are used by various learning algorithms to create n classifiers.

Fig: Standard Stacking Ensemble

1) GA-Stacking:

GA-Stacking is the acronym of Genetic Algorithms for Stacking. GA-Stacking predict how many learning algorithms are necessary to generate base-level classifiers and, for meta classifier what algorithm should be used to generate. The GA-Stacking solution consists on considering them as an optimization problem, which can be solved by applying Genetic Algorithms. The application of Genetic Algorithms to an optimization problem needs, mainly, the study of two aspects: the encoding of the candidate solutions and the definition of the fitness function [6]. The process of codification of the solutions takes place before the execution of the Genetic Algorithms. With respect to the fitness evaluation, it is an iterative process that is carried out in each generation of the Genetic Algorithms on all the individuals (p) of the population (P)[6].

Phase 1: Before the execution of the genetic algorithms the process of codification of the solutions takes place. On all the individual population with respect to the fitness evaluation it is an iterative process that is carried out in each generation of the Genetic algorithms.

Phase 2: Encoding of Candidate Solutions: Binary representation is used in order to represent the candidate solutions or individuals in Genetic Algorithms. The size of the chromosomes can be selected by generating the base level and meta-classifiers using number of algorithms. Also, the maximum number of base-level classifier in the ensemble.

2) Ant Colony Optimization with Stacking:

Considering the outstanding performance of ACO in different applications, we extend the application of ACO in stacking configuration optimization. In an ACO-Stacking construction task, a set of base-level classifier candidates and a set of meta-classifier candidates are given as well as the training sets, the validation sets and the testing set [7]. The base-level classifiers in the set are taken as the “paths” to be selected by the ants. For each iteration, an ant tries to select a path in its route to achieve better performance. Each ant is assigned a certain meta-classifier to combine with the selected “paths” into the “path” package of the ant [7]. A stacking model is configured with the base-level classifiers (“paths” of the ant) and the meta-classifier. This stacking is then trained with the training set(s) and validated with the validation set(s) [7]. If the new “path” package is better than the existing one, it will replace the existing package. Otherwise, the existing “path” package of this ant does not change [7].

At the end, the configuration (the “path” package) of the best ant will be the final configuration of the approach. Finally, this configuration is tested by using the test set. The above process is given in Fig. 1. In the following subsection, the algorithm framework of ACO-Stacking is discussed [7].

Step 1: INPUT
- Datasets: Training Sets, Evaluation Sets
- Building technique for meta-classifiers and base-classifiers

Step 2: Framework for Ant Colony Optimization
- Stacking with Ant Colony Optimization
- Evaluating and Training Stacking
- As a result, best Ant as an output.

Step 3: Testing: . . . . .

Fig 1.1 Ant Colony Optimization-Stacking

IV. PROPOSED METHODOLOGY

In recent years disease predicting using soft computing techniques is most important area of research in data mining. To classify heart diseases, this paper proposes a diagnosis system using cuckoo search optimized rough sets based attribute reduction and fuzzy logic system also We test the CS-based modified stacking algorithm on ten data sets from the UCI Data Repository and show the enhancement in performance over the individual learning algorithms as well as over the typical stacking algorithm. The disease prediction is followed with these steps 1) feature reduction using cuckoo search with rough set theory 2) Disease prediction by means of fuzzy logic system. The first step decreases the computational problem and improves performance of fuzzy logic system. Second step is based on the fuzzy rules and membership functions which categorizes the disease datasets. We have tested this approach on Cleveland, Hungarian, Switzerland heart disease data sets and a real-time diabetes dataset. The experimentation result proves that the proposed algorithm outperforms the existing methods.

A. Feature Reduction Using Cuckoo Search With Rough Set Theory

Fuzzy Logic is a many-valued reason. Fuzzy Logic deals with values which range from complete truth to partial truth, partial false and ultimately to complete false.

Step 1 (Fuzzification): This step translates hard input to fuzzy truth values with help of fuzzy membership functions. They may have any shape. Gaussian functions are used to get better approximation capabilities.

Step 2 (Reasoning by Fuzzy If-then rules): Reasoning is achieved by an inference engine that infers a decision conditional on collective activation degrees of rules in a rule base.

Step 3 (Defuzzification):
With help of a defuzzification function, the result of above reasoning is a fuzzy value which is converted into a corresponding crisp value.
B. Disease prediction using fuzzy logic system

Step 1 (Pre-processing): In this first step of Preprocessing involves missing value handling, normalization, noise removal, and feature selection.

Missing value handling
Due to the presence of noise and missing values in the dataset, the accuracy of any classifier, pre-processing of dataset is a vital step in classification. There are various methods for detection and handling of missing values. For the current dataset, examination of data indicated that missing values have been disguised as zeros. In this case, features with missing values are removed from the dataset.

Normalization
Since different features are measured on different scales, normalization is essential for undistinguishable treatment of several features.

Feature Selection
We have to select only the most useful features from the training dataset for simpler FRM implementation and improved classification accuracy. The features with maximum significance are retained by the system and the rest are rejected.

Noise Removal
Noise denotes to outlier samples in the training dataset. K-means clustering algorithm has been used to detect outliers. During clustering, we initially use all input attributes of the dataset (except those discarded due to missing values). Then we remove the least significant attribute. (Wikipedia, 2013). This procedure is sustained until we achieve clustering using the only two top most informative attributes.

Step 2 (Classifier Rule Base Construction): Initially an FRM rule base is created by means of all mixtures of selected input parameters and equivalent output according to training dataset. In this rule base, number of rules is equal to \(Mn\) where, \(M\) is the no. of membership functions per parameter and \(n\) is the total number of attributes/parameters.

Step 3 (Classifier Optimization): Cuckoo Search has been used to enhance the outputs of fuzzy rules and shape. The search for finest performance is showed by a fitness function that estimates the FRM performance for each combination of fuzzy membership function shapes by means of classifier performance metrics.

C. Combining modified stacking with Cuckoo Optimization:

The stacking ensemble learning using Cuckoo optimization as the meta-learner is shown in figure 2. The training data set is divided into a training subset and a holdout subset. The training subset is further divided into \(n\) subsets using stratified sampling with replacement, creating \(n\) classifiers by using different learning algorithms. The Cuckoo search algorithm is then used for predicting the test set instances also to learn a weight distribution vector that creates the meta-classifier.

Fig Stacking Ensemble Learning using Cuckoo Algorithm

Cuckoo search optimization implements a weight distribution vector. These distribution vectors act as an individual member of the population. Each population member is a vector of weights for each classifier. We select the following operators and parameter values for the Cuckoo search optimization based on some initial set of experiment runs. We select a mutation operator, operator and a standard one-point crossover operator, and. With the help of one value from the vector of weights for an individual is randomly transformed by a small amount. When weights do not add up to 1.0 then they create an invalid vector. After then we normalize the vector by dividing each weight value by the sum of all weights. In utmost cases, the optimum settings would vary with data sets. Instead, our goal is to show the efficiency of this modified stacking with cuckoo search optimization algorithm.

Table shows the observations for algorithms based stacking ensemble optimization techniques.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Observations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stacking</td>
<td>Obtaining optimal configuration of base and meta-level classifiers are problem</td>
</tr>
<tr>
<td>GA-Stacking</td>
<td>Give more importance to speed than accuracy</td>
</tr>
<tr>
<td>ACO Stacking</td>
<td>Better in terms of configuring the system effectively.</td>
</tr>
<tr>
<td>CO-Stacking</td>
<td>Prediction of classifiers are estimated in accurate.</td>
</tr>
</tbody>
</table>

The cuckoo begins by creating a weight distribution vectors. The calculation of each population member is done by estimating the corresponding meta-classifier created by using its weight distribution vector on the holdout subset. The prediction accuracy of that meta-classifier on the holdout subset is calculated to be the fitness of each member. Using the fitness of each population member, the cuckoo is used to select members for the next generation. It then spread over the crossover and mutation operators to create a new generation of weight distribution vectors. The above process is repeated for five thousand generations, and the best weight distribution vector from its final population is selected to create the meta-classifier.

V. RESULTS AND DISCUSSION

The Cleveland, SPECT and Stat log heart disease data sets used for the study were taken from the UCI Data Repository (http://archive.ics.uci.edu/ml/). The standard stacking algorithm was able to improve the prediction accuracy on these heart disease data sets. The modified stacking algorithm with Cuckoo was able to progress on the performance of the standard stacking algorithm. The best improvement in performance was on the Cleveland heart
disease data set where the modified stacking algorithm was able to improve the prediction when compared to the standard stacking algorithm.

Table shows the observation of accuracy based on the optimization algorithms with simulated ensemble techniques.

<table>
<thead>
<tr>
<th>DATASET</th>
<th>STACKING</th>
<th>GA-STACKING</th>
<th>ACO-STACKING</th>
<th>CO-STACKING</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cleveland</td>
<td>85.07</td>
<td>97.61</td>
<td>97.54</td>
<td>98.65</td>
</tr>
<tr>
<td>SPECT</td>
<td>96.17</td>
<td>95.03</td>
<td>94.03</td>
<td>96.28</td>
</tr>
<tr>
<td>Statlog</td>
<td>83.05</td>
<td>76.78</td>
<td>73.48</td>
<td>84.16</td>
</tr>
<tr>
<td>Indian Dataset</td>
<td>68.05</td>
<td>76.01</td>
<td>75.06</td>
<td>77.12</td>
</tr>
</tbody>
</table>

VI. CONCLUSION

In the earlier research, different stacking techniques based on Genetic Algorithm, Ant Colony Optimization, Particle Swarm Optimization has been presented. On examining and studying several versions of Genetic based Stacking based Stacking, it is understood that optimal configuration of the stacking ensemble differs with the type of classifier which is used as base and meta classifiers. Cuckoo -stacking is more efficient and provides more relative enhancement when compared with Genetic-Stacking and ACO-Stacking Accuracy, Speed, Correctness, Relative improvement are the measures which help to differentiate between the Genetic-stacking, ACO-Stacking and Cuckoo-Stacking In future research work it is vital to optimize the stacking ensemble configuration by using other search algorithms such as firefly etc.

REFERENCES
