DATA ANALYSIS USING BAT ALGORITHM FOR VECTOR MACHINES CLASSIFIER IN R-PROGRAMMING

S.Priyadharisini, Assistant Professor/BCA, K.S.Rangansamy Arts and Science College(Autonomous),Tiruchengode,

Dr.J.K.Kanimozhi, Professor/MCA, Sengunthar Arts & Science College, Tiruchengode

ABSTRACT:

Support Vector Machines (SVM) are considered to be great patterns classification techniques. The process of classifying a model with high classification accuracy counts mainly on tuning Support Vector Machine parameters which are the generalization error parameter and the kernel function parameter. SVM represents supervised machine learning approaches, and it is an excellent classifier built on statistical learning approach. This can be done by optimizing the datasets by using Multi-Layer Perceptron (MLP) Model. The SVM parameters and resulted in low classification accuracy. Most approaches related to solving SVM model selection problem will discretize the continuous value of SVM parameters which will influence its performance.They are new, and much of their relative performance is still using **(**Comparative Performance Analysis of Bat Algorithm**)**, it is used to discover patterns in large datasets and has full applications in the field of statistics. Data mining techniques are devised to address forecasting problems by providing a reliable model with data mining features. The result is expected local search when the results are converging well resulting into significant increase in execution time. Expectedly, BA yields faster convergence to the solution.

Keywords:Support Vector Machines,Multi-Layer Perceptron, Patterns Classification, Bat Algorithm.

INTRODUCTION

Support vector machines (SVMs) are a powerful and widely used classification tool for supervised learning, which are based on the VC dimension and structural risk minimization principle in statistical learning theory. The original SVM algorithm is a linear classifier invented by Vapnik, which constructs two parallel hyperplanes between two classes of examples in a finite-dimensional space and maximizes the distance between the two parallel hyperplanes. Boser et al. recommended a way to create nonlinear classifiers by relating the kernel trick to maximum-margin hyperplanes. These complications affect each other and in turn, will effect on SVM's enactment. No conservative methodology agrees on the advance approximation of optimal values for SVM limits. In present classification work, obtaining good values for these parameters is not easy. It requires either a comprehensive search through the space of hyper variables or an optimization method that searches merely a bounded sub-group of the possible values.

Soft Margin Classification:

Extensive datasets invariably have noisy data points. There will be in general no long separation in the input space. In which case, we will have training examples between the two supporting hyperplanes. We need a way to tolerate noise and outliers, and take into consideration the positions of more training points than just those closest to the boundary.

MLP:

A multilayer perceptron (MLP) is a feed-forward artificial neural network that produces a set of productions from a set of inputs. An MLP is branded by numerous layers of input nodes connected as a focused graph between the input and production layers. MLP uses backpropagation to exercise the network. MLP is an in-depth learning method. Multi-layer perceptron's (MLP) is a popular form of feedforward artificial neural networks with many effective presentations in data classification.

Gestational Breast Cancer

Gestational Breast Cancer is one type of BC that occurs during or after pregnancy and are recognized as one of the most common pregnancy's complication. Therefore, this study aims at applying IEM

on a real clinical dataset to predict the occurrence of Type 2 BC from a gestational Breast Cancer (GBC) that occurs during or after pregnancy in Taiwan. Particle representation we use binary digits for feature representation in applying IEMto the feature selection problem. The bits consist of 0 and one number,corresponding to non-selected and selected features respectively.Each particle is coded as a binary alphabetical string. For instance,particle 10,100 contains five elements where only the first and thethird features are selected.

Accuracy is the most commonly used parameter in classifier performance assessment. In the present study, the classification accuracy is determined through classification rate. Efficiency is used to evaluate classifier performance by defining the total number of proper classifications over the total number of available examples. The classified test points can be divided into the following four categories, which are usually represented in the well-known confusion matrix: true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN).

The column of the matrix denotes the instances in a predicted class, while the row denotes the cases in an actual level. Given the four categories of the confusion matrix, accuracy is defined as the fitness function. With the development of the global tourism,academic tourism research also achieves active development, and the focus of research is the traveldemand forecasting. From the last century, the Qualitative and quantitative tourism demand forecasting methods had been developed rapidly. In a broader sense, the existing mg quantitative methods of forecasting tourism demand can be divided into two categories: non-causality forecasting methods (mainly time-series analysis) and causality forecasting methods (primarily econometric model).

Local Search: The procedure that searches for a better solution by collecting the localized information from each sample point. The pseudo-code of the local search procedure.

Simulated Annealing

The EM algorithm is derived from the attraction-repulsion theory of physics and considered as a category of the population-based algorithm. It differs from both the GA and SA (Simulated Annealing) regarding swapping the information between the individuals of the population, but bears a striking

resemblance to both the Particle-Swarm Optimization (PSO)m and Ant-Colony Optimization (ACO), because the particles affect each other in the population. This reproduction operation involves two parents chosen from a pool of selected chromosomes producing an offspring.

This work uses the order-crossover method to produce new genetic materials for the first portion of the chromosome. For example, if two parent chromosomes are present, then the two selected random points will divide the respective parents into a left, middle, and right portion. This results in the first child inheriting the respective left and right section from the first parent, while the central part is found by the genes in the central part of the first parent in the instruction in which the standards appear in the second parental. Similar events were accepted out for the second child.

The second and the third parts of the chromosome use the two-point limit. EM algorithm providing multiple solutions at each task point, and GA used these solutions during the tour of dual manipulators. , GA selected the solution in such a way that fulfills the objective, where the robot can pass the sequence of points safely and with less cycle time. The answer might not suggest an optimal path for the robot to follow but requires shorter cycle time, primarily when it is resulting from the simultaneous operation of several actuators at the joints. Reconstruction of excellent shape from scanned signals involves the solution of an inverse problem and consists of the calculation of a large number of forwarding problems. Finite Element Method (FEM) can be used to represent the forward process. However, iterative methods using a numerically based forward model are computationally expensive. In this situation the Least Square Support

Vector Machines

Vector Machines (LS-SVM) is an excellent alternative to iterative methods and that they exhibit good generalization capability. The Job Shop Scheduling (JSS) problem is one of themost well-known combinatorial optimization complications. Theproblem is identified to be NP-Hard and has been studied fora long time., we are giving a new hybrid met heuristic.

Guided Local Search

Approach for JSS with the arrangement of the Electromagnetism-like method (EM) and Guided Local Search (GLS) and also we used 2-Opt neighborhood job within GLS. EM is one

of the meta-heuristic optimization approaches that operate in a multi-dimensional resolution space where each solution point is moved by forces that obey the rules of Electromagnetism, and Local Search plays a vital role in it.

Our computational experimentationsspecify that the new EM method is capable of manufacturingdependablyoutstanding results for stimulatingexamples of the JSS. Which has been used in our proposed hybrid EM algorithm and the simple local search of the original EM shows that our method makes better and faster the implementation to find the value of objective function. (We run each algorithm 25 times for each problem, then record the best result and average time.)

Optimal power flow

The aim of the optimal power flow (OPF) problem is tocontrol optimal outputs of generating units and find theoptimal sets of a given power system network to reliably supply the classification load at the lowest likely cost while n sufficient system and creating units constraints. The OPF is a nonlinear, non-convex and non-differentiable optimization problem. This is due to the non-smooth or non-convex characteristics of the cost meaning and physical confines of generating units.

Incremental Continuous Ant Colony Optimization

According to Incremental continuous Ant Colony Optimization (IACO) start with a small archives size, a limit Init Archive Size defines it's to scale. IACO begins with randomly resetting solution archive. This answer archive will be filled with solutions initially generated arbitrarily. Each time a new result is added, it is ready by using the material from the best result in the archives. IACO is used to solve SVM model selection problem. In optimizing discrete variable, incremental mixed inconstant ant colony optimization uses the same method that has been used in ACO.

LITERATURE SURVEY

Optimal Power Flow Problem Considering the Cost, Loss, and Emission by Multi-Objective Electromagnetism-Like Algorithm,Considering simultaneous minimization of several contradictory and non-commensurable objectives, such as transmission line loss and pollutant emissions, concludes to a multi-objective OPF (MOOPF) [1]. Electromagnetism-like Apparatus Algorithm and Least Square Support Vector Mechanism for Estimation

the Defect in Nondestructive Evaluation, Reconstruction of excellent shape from skimmed signals requires the result of an inverse problem and includes the control of a large number of forwarding problems. Finite Element Method (FEM) can be used to represent the forward process. However, iterative methods using a numerically based forward model are computationally expensive [2]. A Novel Approach for Multispectral Satellite Image Classification Based on the Bat,

Satellite images can also be a viable source for investigating the temporal changes in the agricultural activities of a particular area. The crop growth, from sowing to harvesting, can be monitored using these satellite images. The orthorectified and dereferenced satellite images can be used to identify problematic areas, and the size of the area affected [3]. Texture Image Classification Based on Maintenance Vector Machine and Bat Algorithm,Remote sensing has been widely used for gathering information on land-use and land cover. Image arrangement is the critical step for extracting information from digital remote sensor data. Image cataloging is vital for bridging the huge semantic gap between images. Consortium contains a variety of decision-theoretic methods to classify pictures [4]. A Multi-Swarm Bat Algorithm for Global Optimization, in real-world life, numerous glitches can be modeled into global optimization problems, and then they are solved by optimization methods. However, with the increment of a complex of the problem, the old optimization methods are hard to find the satisfactory solutions. Inspired by phenomenon in nature, various nature-inspired met heuristic algorithms [5].

Training neural networks by electromagnetism-like mechanism algorithm for tourism arrivals forecasting, usually, there exists error between the actual outputs and the excepted outputs for each training sample. The neural network training aims to find the appropriate parameters to minimize the error. In our tourism arrivals forecasting problem, there is only one output node because their forecasting indicator is just the number of visitors. So MAPE (mean absolute percentage error) is proper for defining the failure [6]. Using Electromagnetism-Like Algorithm and Genetic Algorithm to Optimize Time of Task Scheduling for Dual Manipulators, proposed an approach that is based on the Artificial Neural Network (ANN) to solve the inverse

kinematics of a six-Degree of Freedom (DOF) robot. Two different configurations for the ANN were constructed. This work uses an extensive data set, while there is no mention of the possibility of multiple solutions for inverse kinematics [7]. A new hybrid Electromagnetism Algorithm for Job Shop Scheduling, The aim Job Shop Scheduling (JSS) problem is one of the greatest important combinatorial optimization problems. The problematic is known to be NP-Hard and has been considered for a long time. In this paper, we are awarding a new hybrid met heuristic method for JSS with the arrangement of the Electromagnetism-like method (EM) and Guided Local Search (GLS), and also we used 2-Opt neighborhood job within GLS [8].

Mission Planning for Electromagnetism Environment Monitors Satellite Based on Heuristic Tabu Search Algorithm, after the previous analysis, task mission planning model is the use of resources to meet the constraint conditions, call the appropriate resources to accomplish more tasks as much as possible at the right time. Where the number of standard more than the work up to the action, nor is the task priority and the largest, instead of using a variety of different scheduling policies for various tasks, get a satisfactory solution [9]. An adapted Bat Algorithm for the Quadratic Obligation Problem, The algorithm has also been successfully applied to a wide range of problems in optimization and planning.However, the present paper presents its first application to the generalized assignment problem. Flutters use a type of sonar, called echolocation to avoid an obstacle, detect prey, and locate their roosting crevices in darkness.

These bats emit a sound and listen for echo when it is returned after colliding with the objects[10]. A Simplified Electromagnetism-like Mechanism Algorithm for Tool Path Planning in 5-Axis Flank Milling, we explain how to encode the tool path planning in EM. Generally, a tool path is represented by the connection between the boundary curves of the ruled surface. Then optimize a series of curve parameter pairs which are the actual variables. However, this encoding can lead to a problem: the cutter has to contact with the boundary curves. This constraint wrongly restricts the search space and may produce an unsatisfactory solution [11]. WSN's Energy-Aware Coverage PreservingOptimization Model

based on Multi-Objective Bat Algorithm, this is repeated till the termination criteria are satisfied. Finally, when all requirements successfully met the best so far, a solution is reported. Bat Algorithm Initialization of the population/solutions of NP bats is set randomly.

Initial locations of all bats (Sensor nodes) are set as initial solutions. The idea of the proposed multi-objective optimization model lies in maximizing the coverage ratio and minimizing the consumed energy for both nodes and cluster heads. The number of the selected cluster heads is not restricted to achieve full coverage with minimum energy consumption [12]. Binary Bat Algorithm for Graph Coloring Problem,The Graph Coloring Problem (GCP) is one of the most interesting, studied and difficult combinatorial optimization problems. The GCP consists in coloring each vertex of a givengraph by using a minimum number of colors called chromatic number so that no two adjacent vertices are painted with the same color. Unfortunately, GCP has been shown to be. Hence, several approaches were developed to handle this problem that we can classify into three classes; exact approaches, heuristic approaches and met heuristics approach. There exist quite a few specific strategies for this problem [13]. Adaptive & Discrete Real Bat Algorithms for Route Search Optimization of Graph-Based Road Network, Bat Algorithm (BA) has lately emerged as an efficient nature enthused meta-heuristics due to its added parameters and searching features discrete domain problems.

Here the BA has used the well-known three variable dependent Weibull Cumulative Distribution Purpose as Weibull Coded Binary Bat Algorithm (WCBBA), another as Real Bat Algorithm (RBA) and the third as the hybrid of the two, for search procedurewith its modeling according to a road network organization system where it is being tried to augment the travel route and produce a vehicle load balancing structure of the network through optimized path establishing [14]. Unsupervised Feature Selection Using Binary Bat Algorithm,Feature selection is choosing a subset of optimal features.

Feature selection is being used in high dimensional data discount, and it is being used in several applications like medical, image processing, text mining, etc. Several methods were introduced for unsupervised feature selection. Among those methods, some are based on filter

approach, and some are based on wrapper approach [15]. Optimal assignment and sizing of DGs at many load circumstances using Shuffled Bat algorithm, the power industry has seen substantial changes in the electrical distribution system mostly due to the implementation of smart grid technology and the incremental infiltration of distributed generation. DGs are positioned near to the load center, and the addition of DG with the grid has been given greater attention due to following advantages such as system loss discount, development in the voltage profile, system constancy and power excellence of the distribution system [16].

A CBPSOGSA-SVM hybrid system for classification**,** In order to solve the problems and improve the performance of SVM for classification, a new proposed chaos-optimized binary version of PSOGSA which is the hybrid of particle swarm optimization (PSO) and gravitational search algorithm (GSA), termed CBPSOGSA, is developed in this paper to build a novel hybrid system (CBPSOGSA-SVM) for classification [17**].** A new brain-robot interface system based on SVM-PSO classifier For SVM-PSO classifier, the EEG signals are collected from two databases: PhysioNet and BCI Competition III, then features including Power Spectral Density (PSD) and wavelet parameters are used as the input of the classifier. By comparing the results of the SVM and SVM-PSO classifiers, is concluded that performance of the classifier in terms of accuracy is increased through PSO algorithm [18].

A detecting method for malicious mobile application based on incremental SVM**,** Due to the rapid growth of Android malicious application samples, traditional detection methods need to spend a lot of time for training, a detecting method for malicious mobile application based on incremental SVM was proposed to achieve additional learning of the detection system. The technique used the SVM as the classification and training algorithm and extracted sensitive permissions and APIs as application characteristics [19]. Parallelized variable selection and modeling based on prediction**,** Lack of safety and efficacy are the two primary reasons for the failures of drug candidates in drug discovery and development. These failures can be addressed using reliable and readily applicable predictive ADMET models at early discovery stage as they rationalize experimental observations. Structure-activity relationships models are

one such class of predictive models that are typically generated from a) in vitro (or) in vivo experimental data b) a large number of structural features that are derived from drug candidate's chemical structure [20].

IMPLEMENTATION OF PROPOSED SYSTEM:

Optimization problem finding the best solution from all feasible solutions. Many meta-heuristic algorithms have been developed in the literature including that of Simulated Annealing (SA, Ant Colony Optimization (ACO to name a few. The fact that they are new and much of their relative performance is still using **(**Comparative Performance Analysis of Bat Algorithm**)**,itis used to discover patterns in large datasets and has broad application s in the field of statistics. Data mining techniques are devised to address forecasting problems by providing a reliable model with data mining features. We use the auto-regressive integrated moving average model to predict the market trends. Optimization data selection problem solving using Support Vector Machines (SVM) technic are considered to be great patterns classification techniques. The process of classifying a model with high classification accuracy counts mainly on tuning Support Vector Machine parameters which are the generalization error parameter and the kernel function parameter. SVM represents supervised machine learning approaches, and it is an excellent classifier built on statistical learning approach.

The central concept of SVM is to obtain the Optimal Separating Hyper-plane between positive and negative samples. This can be done by optimizing the datasets by using Multi-Layer Perceptron (MLP) Model. The SVM parameters and resulted in low classification accuracy. Most approaches related to solving SVM model selection problem will discretize the continuous value of SVM parameters which will influence its performance.

MAIN CONTRIBUTION

1. The algorithm has used the BA to optimize SVM classifier parameters.
2. The error signal produced actuates a control mechanism of the learning algorithm.
3. hey are not as favored for large-scale data mining as for pattern recognition or machine learning because the training complexity of

SVMs is highly dependent on the size of data set.

4. The organization accuracy of the built SVM classifier is utilized to direct the bringing up-to-date of solution archives and pheromone table.

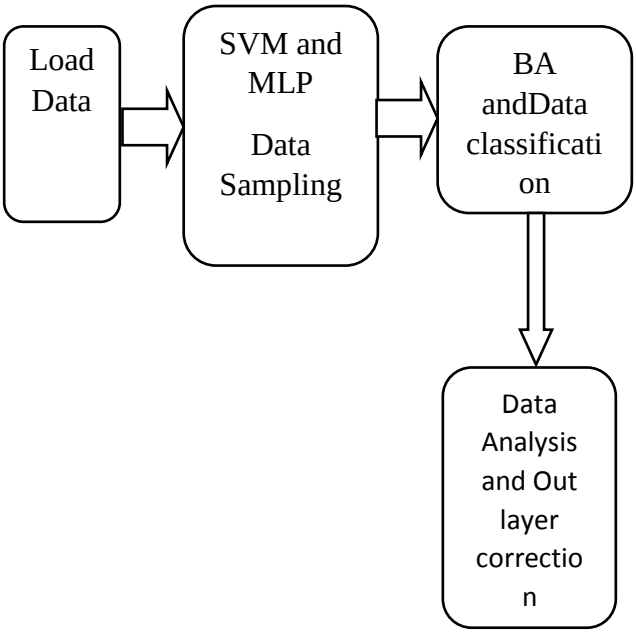
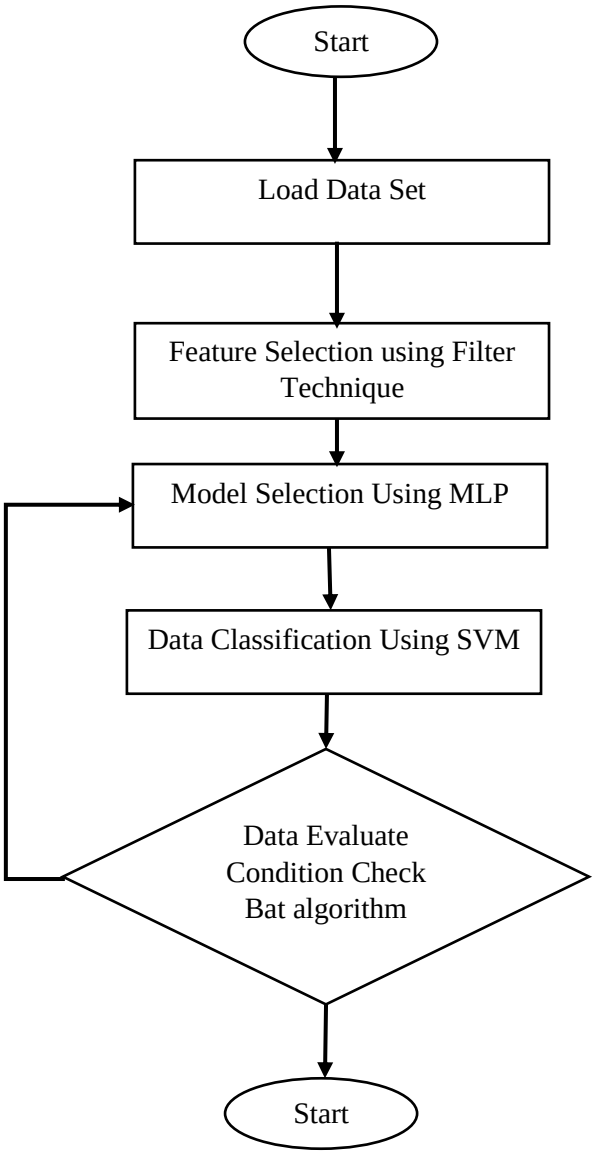


Fig: Data classification using SVM and BA algorithms

The above show in the figure first load data set using R-programming used and SVM and MLP is uploading dataset from select and prediction method applied to classified data's from the dataset, Bat algorithm is All bats are using echolocation to calculate distance intelligently. They know the difference between food/prey and the surrounding environment background in a magical way. The algorithm has been built on the assumption that the bat is able to find its prey in complete darkness. Finally, data classified data out layer correction create in a graph.

DATA FLOW DIAGRAM



 ISBN : 978-93-87793-75-0

PARTS OF THE PROPOSED SYSTEM

1. **Support Vector Machines (SVM)**
2. **Multi-Layer Perceptron (MLP)**
3. **Dataset Classification**
4. **Outlier handling**

Support Vector Machines (SVM)

Support vector machines (SVMs) are a powerful and widely used classification tool for supervised learning, which are based on the VC dimension and structural risk minimization principle in statistical learning theory. The original SVM algorithm is a linear classifier invented by Vapnik, which constructs two parallel hyperplanes between two classes of examples in a finite-dimensional space and maximizes the distance between the two parallel hyperplanes. Boser et al. free as a way to generate nonlinear classifiers by smearing the kernel trick to maximum-margin hyperplanes. These glitches affect each additional and in turn, will effect on SVM's performance. There is no conservative methodology that accepts the advance approximation of optimal values for SVM parameters. In present classification work, obtaining good values for these parameters using following steps.

Step 1: Logistic Regression works when we have data which is linearly separable.

Step 2: linear separating boundary in this case.

Step 3: divide this data into the current dimensions (1D) but can there be any boundary.

Step 4: multiple possible such hyperplanes which divide the data Consider point.

```
> library(class)
> predictions <- knn(train = trainingSet, cl = trainingOutcomes, k =
+                    test = testSet)
> predictions
  [1] benign    malignant benign    benign    benign
  [6] benign    benign    benign    benign    benign
 [11] benign    benign    benign    benign    malignant
 [16] benign    benign    benign    benign    benign
 [21] benign    benign    malignant malignant benign
 [26] benign    benign    malignant benign    benign
 [31] malignant malignant benign    benign    benign
 [36] benign    benign    benign    malignant benign
 [41] benign    benign    benign    benign    benign
 [46] benign    benign    benign    benign    benign
 [51] benign    benign    benign    benign    malignant
 [56] benign    benign    malignant benign    benign
 [61] benign    benign    benign    benign    benign
 [66] benign    benign    benign    benign    benign
 [71] benign    benign    benign    malignant benign
 [76] benign    malignant malignant malignant malignant
 [81] benign    benign    malignant benign    benign
 [86] benign    benign    benign    benign    malignant
 [91] malignant benign    benign    benign    malignant
 [96] benign    malignant benign    malignant malignant
[101] malignant benign    malignant benign    benign
[106] benign    benign    benign    benign    benign
[111] benign    malignant malignant malignant benign
[116] benign    malignant benign    malignant malignant
[121] malignant benign    benign    benign    benign
[126] benign    benign    benign    benign    benign
[131] benign    benign    benign    malignant benign
[136] benign    benign    benign    benign    benign
[141] malignant benign    benign    malignant benign
[146] benign    benign    benign    benign    benign
[151] benign    benign    benign    benign    benign
[156] malignant benign    benign    benign    benign
[161] benign    benign    benign    benign    benign
[166] malignant benign    benign    benign    benign
[171] benign    benign    benign    benign    benign
[176] malignant malignant malignant benign    benign
[181] benign    benign    benign    benign    benign
[186] benign    benign    malignant malignant benign
[191] benign    benign    benign    benign    benign
[196] benign    benign    benign    malignant benign
[201] benign    benign    benign    malignant malignant
Levels: benign malignant
```

The above showed in the figure is represented by prediction methods it is predicted by the cancer dataset in to be dices affected levels benign is served to start affected areas, and malignant is rendered non-affected areas.

```
> summary(breastc)
 sample code number clump thickness  uniformity of cell size
 Min.   :   61634   Min.   : 1.000   Min.   : 1.000
 1st Qu.:  870688   1st Qu.: 2.000   1st Qu.: 1.000
 Median : 1171710   Median : 4.000   Median : 1.000
 Mean   : 1071704   Mean   : 4.418   Mean   : 3.134
 3rd Qu.: 1238298   3rd Qu.: 6.000   3rd Qu.: 5.000
 Max.   :13454352   Max.   :10.000   Max.   :10.000
 uniformity of cell shape marginal adhension
 Min.   : 1.000           Min.   : 1.000
 1st Qu.: 1.000           1st Qu.: 1.000
 Median : 1.000           Median : 1.000
 Mean   : 3.207           Mean   : 2.807
 3rd Qu.: 5.000           3rd Qu.: 4.000
 Max.   :10.000           Max.   :10.000
 single epithelial bare nuclei        bland chromatin
 Min.   : 1.000    Length:699         Min.   : 1.000
 1st Qu.: 2.000    Class :character   1st Qu.: 2.000
 Median : 2.000    Mode  :character   Median : 3.000
 Mean   : 3.216                       Mean   : 3.438
 3rd Qu.: 4.000                       3rd Qu.: 5.000
 Max.   :10.000                       Max.   :10.000
 normal nucleoli     mitoses           class
 Min.   : 1.000   Min.   : 1.000   Min.   :2.00
 1st Qu.: 1.000   1st Qu.: 1.000   1st Qu.:2.00
 Median : 1.000   Median : 1.000   Median :2.00
 Mean   : 2.867   Mean   : 1.589   Mean   :2.69
 3rd Qu.: 4.000   3rd Qu.: 1.000   3rd Qu.:4.00
 Max.   :10.000   Max.   :10.000   Max.   :4.00
> hist(breastc$`clump thickness`)
> hist(breastc$`class`)
> mean(breastc$`clump thickness`)
[1] 4.41774
> max(breastc$`clump thickness`)
[1] 10
> mean.Date(breastc$class)
[1] "1970-01-03"
> mean(breastc[,2])
[1] NA
```

The above showed in the figure is represented min keywords are represented minimum affected areas, Median is middle level, Max keyword is represented maximum affected areas it will be classified varies of Multi levels layers classification used for MLP methods.

Multi-Layer Perceptron (MLP)

The learning process of MLP network is based on the data samples, composed of the N-dimensional input vector x and the M-dimensional desired output vector d, called destination. By processing the input vector x, the MLP produces the output signal vector y(x,w), where w is the vector of adapted weights. The error signal generated actuates a control mechanism of the learning algorithm. The remedialalterations are intended to make the output signal y_k (k=1, 2, M) to the desired response d_k in step by step manner.

Dataset Classification:

Margin maximization: The classification boundary functions of SVMs maximize the margin, which in machine learning theory, corresponds to optimizing the generalization performance given a set of training data.

Nonlinear transformation of the feature space using the kernel trick: SVMs handle a nonlinear classification efficiently using the kernel trick which implicitly transforms the input space into another high dimensional feature space.

The success of SVMs in machine learning naturally leads to its possible extension to the classification or regression problems for mining a vast amount of data. They are not as favored for large-scale data mining as for pattern recognition or machine learning because the training complexity of SVMs is highly dependent on the size of the dataset

Outlier handling:

After the construction of a CF tree, the leaf entries that contains far fewer data points than average are considered to be

outliers. A low setting of outlier threshold can increase the classification performance of CB-SVM especially when the number of data is relatively large compared to the number of dimensions and the type of boundary functions are simple (which is related to having a low VC dimension in machine learning theory) because the non-trivial amount of noise in the training data which may not be separable by the simple boundary function prevents the SVM boundary from converging in the quadratic programming.

ALGORITHMS:

Bat Algorithm

Begin

Step 1: Initialization. Set the generation counter t= 1; randomly initialize the population; define loudness Ai, pulse frequency Q, and the initial velocities Vi (I = 1,2, · · ·, NP); set pulse rate Rio

Step 2: While t<MaxGeneration do

Generate new solutions by adjusting frequency,

and updating velocities and locations

if (rand>r,) then

Select a solution among the best solutions

Generate a local solution around the selected best solution

endif

Generate a new solution by flying randomly

if (rand<Ai &j{x,)<j(x*)) then

Accept the new solutions

Increase r, and reduce Ai

end if

Rank the bats and find the current best x*

t = t+ 1;

Step 3: end while

End

Pseudo-code for the hybridized BA is shown below.

Objective function $f(x) = (x1, x2, x3, ...xD)$

Initialize bat population

$Xi,j(i = 1, 2, 3, ...N)(j = 1, 2, 3, ...D)$

Define pulse frequency *fi* at *xi*

Initialize pulse rates *ri* and loudness *Ai*

Iteration counter (*t*) set to 0

while *t* < *max iterations*

if *t* is even

Perform bat search procedure using Eq. (6) - (8)

else

Apply onlooker search procedure by Eq. (18)

end if

if (*rand* > *ri*)

Select random solution among the best solutions

Generate a local solution using random walk Eq. (9)
end if
if ($rand < Ai$ and $f(xi) < f(x*)$)
Accept new solutions
Increase ri and reduce Ai
end if
Rand all bats and find current best $x*$
end while

SUPPORT VECTOR MACHINE (SVM) PROCESSING:

IACOM is an ACOR algorithm with a solution archive whose size increases over time. This modification is based on the incremental social learning framework. A parameter Growth controls the rate at which the archive grows. Fast growth rates encourage search diversification while slow ones encourage intensification. In IACOM the optimization process begins with a small archive, a parameter in it Archive Size denes its size. A new solution is added to it every Growth iterations until a maximum archive size.

Continuous ant colony optimization (introduced by and it used Probability Density Function (PDF) instead of Discrete Probability Distribution to determine the direction that an ant should follow; Gaussian function, an example of PDF is one of the most popular as it uses a straightforward manner for data sampling. For each built solution, a density function is generated from a set of k solutions that the technique preserves at all times. In order to maintain this set, the collection is filled with nonsystematic solutions at the beginning. This is similar to initializing pheromone value in a discrete ACO approach. Then, at each loop, the group of created m solutions is appended to the set, and the equivalent number of worst solutions is deleted from the collection to preserve just the best k solutions of the k + m solutions that are available. This work is similar to pheromone modification in discrete ACO.

The goal is to influence the searching procedure to gain the best solution. Pheromone information is kept in a table when ACO for discrete combinatorial optimization is used. During each loop, when selecting a component to be appended to the current partial solution, an ant utilizes part of the values from that table as a discrete probability distribution.

MLP (Multilayer Propagation):

The performance of multilayer perceptron is determined by the number of hidden layers and by the name of neurons in each hidden layer these are the network

design parameters that are adjusted. The correct classification function is introduced as the ratio of a number of correctly classified inputs to the whole amount of data.

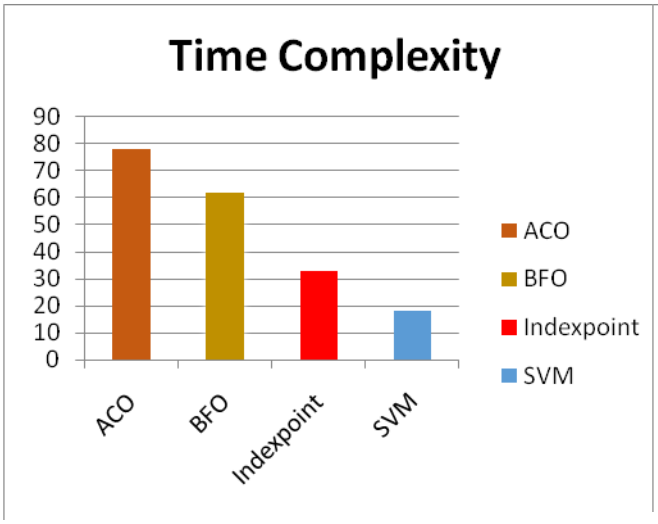
Multilayer perceptron's with 1 and two concealed layers are examined. The process of adjusting the number of neurons in hidden coatings is planned as a grid search (see Appendix). With each mixture of numbers of neurons in the hidden layers, the multilayer perceptron is trained on the train set, the value of correct classification meaning for the train set is stored. The endorsement set is used for standard early stopping procedure, the amount of right organization function for the authentication set is stored as well.

The number of neurons that safeguards the best simplification is chosen. The training and simulation of the designated model of multilayer perceptron are done on joined training and authentication sets, the value of true cataloging function is designed.

RESULT AND DISCUSSION:

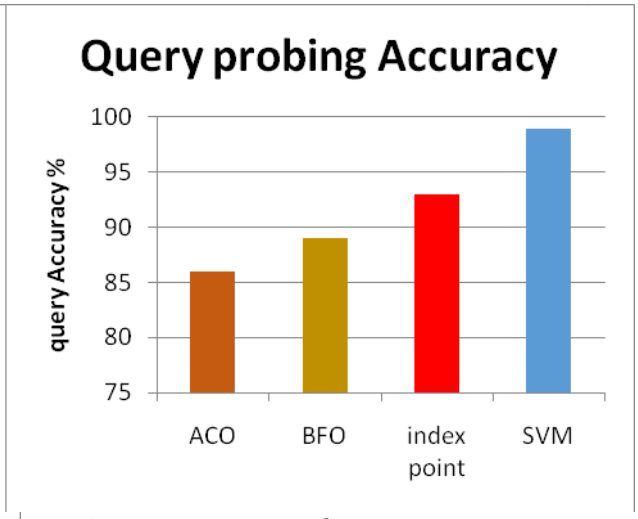
We proposed an SVM method for feature selection by uniting bat algorithm with the adjacentnational classifier and MLP as the local search. We applied the proposed bat method on datasets collected from the source and eight public microarray dataset to verify its presentation on datasets with many characteristics, sizes, and features. Besides, we associated the bat's performance with those of other nine common feature selection methods. We performed a non-parametric regular comparison. The results show that the average correctness and Kappa index of the proposed bat is outdone than those of the other methods. We also recognized that the IEM mean correctness and Kappa index are suggestivelydissimilar from those of the other ways.

At last, some theoretic analysis should be done in order to prove conference of the method. This analysis can also suggestivelyprogress the dependability of the procedure when solving real-world submission problem.



Graph: Comparison of time complexity

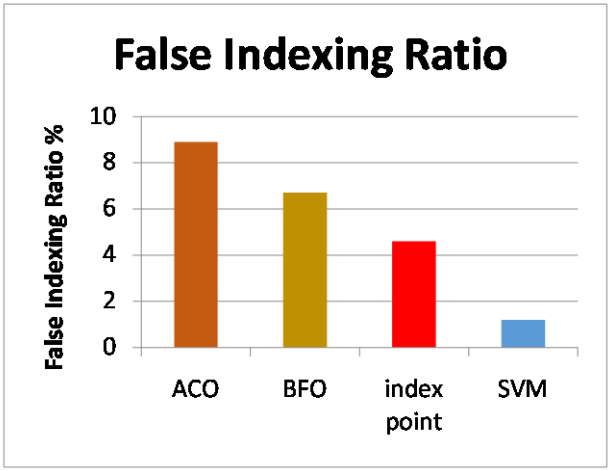
The above chart shows the time complexity of scene produced by different techniques, and it shows clearly that the proposed methods have produced less time complexity than other methods. In graph-based event reference, a graph is created to express the exchanges of entities in a recommender system, where objects are epitomized as location and connections between them are epitomized as edges. Based on the graph model, the recommendation task is then converted into a proximity problem. In the proposed system, the time complexity in clustering to be low likened to another method.



Graph: Comparison of query accuracy

The above Graph shows the similar result of query accuracy produced by different methods, and it shows unmistakably that the proposed plan has formed more query accuracy than other techniques.

We use to represent the number of classifications for the data sets and to represent the number of index point temporal for the out layer. It is used to describe the change between index point and of the same method. The value can determine the likelihood that the two ways are relayed or not. The advanced the value, the subordinate the likelihood that the two affairs are related, and the lower the likelihood that the system is used as a stepping-stone. Where associateBAT must be higher accuracy, comparedto a further method.



Graph: Comparison of false indexing ratio

The above chart shows the mean indexing ratio produced by different algorithms, and it shows clearly that the proposed method has produced less false indexing than other techniques.

ACO is constructed in a similar way that has been done in, which consists of all the protein domain types of yeast proteins in the online model. An edge is introduced to connect two domain types if they coexist in one protein. After constructing ACO, we implement different steps of random walks in the clustering, in order to utilize functional information from different level neighbors of proteins, terms and of domains to predict protein functions. To formally define our method, some variables are introduced. So false indexing occurs during process time. The false indexing occurs higher in EMI when compare to trajectory-data,index point, and BAT. But in proposed system BAT could be low false indexing.

CONCLUSION

This paper proposes a new method called SVM (Clustering- Based SVM) that integrates a scalable clustering method with an SVM method and effectively runs SVMs for massive datasets. The existing SVMs are not feasible to run such datasets due to their high complexity on the data size or frequent accesses on the large data sets causing expensive I/O operations. Continuous Comparative Performance Analysis of Bat Algorithm as an extension of the Comparative Performance Analysis of Bat Algorithm offers the opportunity to deal with constant optimization problems. On diabetes and glass datasets accuracy is almost equal, and Multilayer Perceptron (MLP) is a slightly better algorithm. Thus we found that Multilayer Perceptron is a better algorithm in most of the cases. Other variants of SVM such as least square SVM can also be used to solve classification problems. Design a mechanism for using many kernel functions and selecting the most successful kernel function that gives the best classification accuracy.

REFERENCES

1) JieGeng, Zhenjiang Miao, and CencenZhong, "Skin detection with illumination adaptation in single image," in IEEE International Conferenee on Multimedia and Expo, 2011, pp. 1- 6.
2) Chen-Chiung Hsieh, Dung-Hua Liou, and Meng-Kai Jiang, "Fast enhanced face-based adaptive skin color model," in International Conferenee on Image Proeessing and Pattern Reeognitioninlndustrial Engineering, 2010, pp. 782027-782027.
3) PratheepanYogarajah, Joan Condell, Kevin Curran, Paul McKevitt, and Abbas Cheddad, "A dynamic threshold approach for skin tone detection in colour images," International Journal of Biometries, vol. 4, no. 1, pp. 38-55, 2011.
4) Wei Ren Tan, Chee Seng Chan, PratheepanYogarajah, and Joan Condell, "A fusion approach for efficient human skin detection," IEEE Transaetions on Industriallnformaties, vol. 8, no. 1, pp. 138- 147, 2012.
5) Insung Hwang, Sang Hwa Lee, Byungseok Min, and Nam Ik Cho, "Luminance adapted skin color modeling for the robust detection of skin areas," in IEEE International Conferenee on Image Proeessing, 2013, pp. 2622- 2625.
6) GeervaniKoneti, ”Parallelized variable selection and modeling based on prediction” algorithm on GPU for feature selection and ADMET model 2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) Year: 2017 Pages: 2268 – 2269.
7) Qusay H. Mahmoud, ImtiazUllah, "A filter-based feature selection model for anomaly-based intrusion detection systems" et.all., IEEE 2017
8) MahdiyehHajibabzadeh , ParvizShahabi "A new brain-robot interface system based on SVM-PSO classifier" et.all IEEE 2017
9) Dongsheng Zhao, Yanjun Li “A CBPSOGSA-SVM hybrid system for classification” et.all IEEE 2018
10) Michael J. Waring, et al, “An analysis of the attrition of drug candidates from four major pharmaceutical companies”, Nat. Rev., Drug Discov., 19 June 2015, 1–12, doi:10.1038/nrd4609.