IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-7, Issue-8, Aug 2019 E-ISSN: 2347-2693

Particle Swarm Optimization based Feature Selection with Evolutionary Outlay-Aware Deep Belief Network Classifier (PSO-EOA-DBNC) for High Dimensional Datasets

M. Praveena[1*], V. Jaiganesh[2]

[1]Dept. of Computer Science, Dr. SNS Rajalakshmi College of Arts and Science, Coimbatore, India
[2]Dept. of Computer Science, Dr. NGP Arts and Science College, Coimbatore, India

Corresponding Author: praveenamarannan@gmail.com

DOI: https://doi.org/10.26438/ijcse/v7i8.6169 | Available online at: www.ijcseonline.org

Accepted: 09/Aug/2019, Published: 31/Aug/2019

Abstract- Data mining research extends its wings to several domains and classification is one of the thrust areas for researchers. The curse of dimensionality is reduced by many optimization techniques and machine learning algorithms. In this research work, a particle swarm optimization based feature selection method is employed to deal with the curse of dimensionality. The PSO algorithm makes use of the fitness function that is obtained from the evolutionary outlay aware deep belief network which conducts classification. 20 datasets are taken for evaluating the conductance of the PSO – EOA – DBNC in terms of classification accuracy and elapsed time. From the results it is significant to notice that PSO-EOA-DBNC out conducts than that of other classifiers.

***Keywords*:** data mining, feature selection, particle swarm optimization, deep belief network, evolutionary algorithm.

I. INTRODUCTION

Data mining is one among the dynamic research territories in the areas of computer science research just as data innovation. Amid the previous two decades there is a learning data disclosure movement helps the data mining to haul out concealed data from the dataset, there is a mammoth measure of machine learning algorithms to be fused for conducting data mining assignments. Generally supervised machine learning algorithms enlarge driving ramifications in the research field of data mining. Machine learning instantly named as ML is a kind of artificial intelligence (AI) that make accessible computers with the ability to be prepared without being straightforwardly customized. ML learning centres on the spreading out of computer programs that can adjust just as change at the purpose of time revealed to brand new data. ML algorithms are extensively characterized into three classifications specifically supervised learning, unsupervised learning and support learning. The progression of machine learning is relating to that of data mining [1]. The two data mining and machine learning look at just as inquiry from start to finish data to appear for examples. Then again, in propensity to separating data for human information similar to the case in data mining applications, machine learning utilizes data to perceive designs in data and change program activities from now on.

Supervised machine learning is the assignment of deduction an importance from named preparing data that comprises of a gathering of preparing precedents. To the extent supervised learning is concerned, each occurrence is a prop encases an information object (which is typically a vector amount) and a fundamental yield esteem (may likewise be alluded as supervisory flag). At first, the supervised learning algorithm does the investigation task from the preparation data and fabricates a needy capacity, for mapping new precedents. An ideal setting potentially makes conceivable the algorithm to accurately decide the class marks for secured cases and the equivalent requires

The supervised learning algorithm to make fewer complexes from the preparation data to canvassed circumstances in an "adjusted" way. The supervised algorithms/techniques are most likely utilized in an assortment of utilization regions that incorporate promoting, money, fabricating, testing, securities exchange prediction, etc. This research work points in proposing PSO based Feature Selection with Evolutionary Outlay-Aware Deep Belief Network (PSO-EOA-DBNC) Classification for High Dimensional Datasets. The point of PSO-EOA-DBNC is to improve the prediction accuracy and furthermore to diminish the time taken for classification. This research article is drafted in the following manner. This section introduces the work. Section 2 gives some of the related works. Section 3 portrays the proposed work. Section 4 presents results and discussions. Section 5 gives the concluding remarks.

II. RELATED WORKS

Numerous classification techniques have been developed, such as artificial neural networks [2], Bayes classifiers [3], and support vector machines (SVM). Classification algorithms often have difficulty dealing with data sets that include a large number of features, which greatly increase the temporal and spatial complexity. Many of the features in an input data set are irrelevant or redundant and should therefore be eliminated. Feature selection is the process of identifying the subset of features that would allow the classifier to conduct most effectively. Numerous previous studies [4–6] have coupled classification algorithms with feature selection methods based on global search methods, such as evolutionary algorithms. Evolutionary computation [7] is a universal meta-heuristic algorithm used to resolve optimization problems. Certain classification methods are inspired by modelling biological behaviour namely genetic algorithms [8] and particle swarm algorithms [9]. It is noteworthy that swarm intelligence employs population-based searches for finding / obtaining solutions to optimization research.

Recently, deep learning [10] has been one of the hot topics in machine learning, pattern recognition, feature extraction and data mining [11]. It is to be noted that layer-by-layer abstraction of thousands of neurons in the brain is crucial [12-16] and are used in machine learning application domains.

III. PROPOSED WORK

3.1. Particle Swarm Optimization for Feature Selection

PSO is a meta-heuristic search procedure that simulates the group of bird's where about to locate the food. Each particle in the swarm denotes a candidate solution (in this work it is set of features) that flies throughout the multidimensional search space. A particle makes use of the best position found by itself and that of its neighbours to budge for an optimum solution. The conductance of each particle is calculated based on the specific fitness function.

At a point of time when the search space is *D*-dimensional and there are *m* particles in the swarm. Each particle is located at position $X_i = \lfloor x_{i1}, x_{i2}, ..., x_{iD} \rfloor$ with velocity $V_i = \lfloor v_{i1}, v_{i2}, ..., v_{iD} \rfloor$, where $i = 1, 2, ..., m$. In the PSO algorithm, each particle moves towards its own best position $(pbest)$ denoted as $Pbest_i = \lfloor pbest_{i1}, pbest_{ii2}, ..., pbest_{iD} \rfloor$ and the best position of the whole swarm $(gbest)$ denoted as $Gbest = [gbest_1, gbest_2, ... gbest_D]$. Every particle modifies its position based on its own velocity and the same is created in random fashion in order to obtain the $pbest$ and $gbest$ positions. As far as every particle in the optimization is concerned, i and dimension s, the updated fresh velocity v_{is} and position x_{is} is computed by making use of the equations (1) and (2):

$$v_{is}^{t} = w v_{is}^{t-1} + c_1 b_1 \left(pbest_{is}^{t-1} - x_{is}^{t-1}\right) + c_2 b_2 \left(gbest_s^{t-1} - x_s^{t-1}\right) \quad \dots (1)$$

$$x_{is}^{t} = x_{is}^{t-1} + v_{is}^{t} \quad \dots (2)$$

Where t is the iteration number. The inertial weight w is used to control the velocity and balance of the exploration and exploitation abilities of algorithm. A large value of w keeps particles at high velocity and prevents them from becoming trapped in the local optima.

A little value of w arranges the particles at lesser velocity and also persuades to deed the identical quest zone. The constants c_1 and c_2 are called acceleration coefficients and are used to regulate the PSO particles to mobilize nearer to the positions of $pbest$ or $gbest$. The b_1 and b_2 are independent random numbers uniformly distributed between 0 and 1. The termination criterion of the PSO algorithm includes the maximum number of generations, the designated value of $pbest$, or no further improvement in $pbest$. A PSO based feature selection is presented in this work that determines the importance of features by ranking them and uses the information to improve the search ability of algorithm.

3.1.1. Measuring the fitness

It is well known that feature selection is the choice of a small number of relevant features to obtain similar or even better classification conductance than that of the use of all features. Thus, in this phase of research two main conflicting objectives are taken into account that denotes the classification conductance and the number of features, both objectives should be lessened. The first objective, classification conductance, is error rate, which is calculated via equation (3) for i-th particle:

$$Error\_rate_i = (FP + FN) / (FP + FN + TP + TN) \quad \dots (3)$$

TP, FP, TN and FN refer to true positives, false positives, true negatives, and false negatives of EOA-DBN classifier respectively.

The second conflicting objective considers the number of selected features. Small number of features reduces the computational cost yet increases the error rate. As normalized values of the objectives provide a uniform search of the problem space, second objective should be calculated according to the equation (4) for i-th particle:

$$Feature\_rate_i = f_i / D,\ f_i = \sum_{j=1}^{D} Z_{i\,j}(t) \ldots (4)$$

Where, D is the total number of features. In this fashion, both objectives are normalized in [0, 1]. External store is used to store the non-dominated solutions found during the search. After calculating the fitness function, the non-dominated solutions are extracted, and the store set is updated.

3.1.2. Feature Ranking and Selection

The obtained features by the store members are gaining its momentum when compared with rest of the features. Consequently, the store can rank the features as mathematically modelled by equation (5).

$$FR = \sum_{k=1}^{|A|} Z_k(t),\ Z_k(t) \in A \ldots (5)$$

Set A denotes the store and $Z_k(t)$ is a decoded member of store. In this way, FR is a vector with D dimensions as $FR = [r_1, r_2, ..., r_D]$ that, r_j, $j = 1, 2, ..., D$, determines the rank of j-th feature. The minimum value of r_j is zero when the j-th feature is not selected by any member of store. The maximum value of r_j is equal to the length of the store when the j-th feature is selected by all members of store.

3.2. Background of DBN

Every layer of hidden variables attempts to learn for signifying the features that confine higher order correlations in the original input data. At the point of time when DBNs is applied to the data classification research, feature vectors from records are used to set the states of the visible variables of the lower layer of the DBN. The DBN is then trained to produce a prospect distribution over the possible labels of the data based on subsequent prospect distribution of the records.

Suppose a data set $S = \{\{x_1, y_1\}, \{x_2, y_2\}, ..., \{x_N, y_N\}\}$ contains a total number of N record pairs $\{x_n, y_n\}$, where x_n the nth data is sample, y_n is the corresponding nth target label. Assume a DBN consists of H hidden layers and the parameters of each layer $i \in \{1, ..., H\}$ by $\theta_i = \{W_i, b_i\}$. Given an input record x from the dataset, the DBN with H hidden layer(s) presents a complex feature mapping function. After feature transformation, soft-max layer serves as the output layer of DBN to conduct classification predictions as parameterized by $\theta_s = \{W_s, b_s\}$. Suppose there are K neurons in the soft-max layer, where the j th neuron is responsible for estimating the prediction prospect of class j, given input of x_H which is the output of the previous layer and associated with weights $W_s^{(j)}$ and bias $b_s^{(j)}$

$$P(y = j \mid x) = \frac{\exp\left(b_s^{(j)} + x_H^T W_s^{(j)}\right)}{\sum_{k=1}^{K} \exp\left(b_s^{(k)} + x_H^T W_s^{(k)}\right)} \ldots (6)$$

Where x_H is the output of the previous layer. Based on the prospect assessment, the trained DBN classifier provides a prediction as

$$f(x) = \arg\max_{1 \le j \le K} P(y = j \mid x) \ldots (7)$$

In practice, the parameters $\{\theta_1, \theta_2, ..., \theta_H, \theta_s\}$ of DBN are massively optimized by statistic gradient descent with respect to the negative log-likelihood loss over the training dataset.

3.3. Outlay-Aware Deep Belief Network

Outlay-aware learning attracts researchers since its working mechanism is to lessen the overall outlay on the training dataset. It is presumed that the aggregate count of existing class labels is denoted as K. Sample data is denoted as x, the outlay of misclassifying x as class j when x actually belongs to class I is denoted as $C_{i,j}$. In addition, $C_{i,j} = 0$, when i = j, that denotes the outlay for precise classification is said to be 0.

Known is the misclassification outlays denoted as $C_{i,j}$, a record need to be grouped into the class that has the least predictable outlay. Based on the decision theory, the Verdict statute lessening the expectation outlay $R(i \mid x)$ of classifying an input vector x into class i can be expressed as

$$R(i \mid x) = \sum_{j=1,\, j \ne i}^{K} P(j \mid x) C_{i,j} \ldots (8)$$

where $P(j \mid x)$ is the subsequent prospect assessment of classifying a record into class j. Given the prior prospect $P(x_n)$, the general Verdict statute denotes which action to take for each record x_n, thus the overall risk R is

$$R = \sum_{n=1}^{N} \sum_{i=1}^{K} R(i \mid x_n) P(x_n) \ldots (9)$$

According to the Bayes decision theory, an ideal classifier will give a decision by computing the Expectancy jeopardy of classifying an input to each class and predicts the label that reaches the minimum overall Expectancy jeopardy. Misclassification outlays denote the penalties for classification errors. In outlay-aware learning, all misclassification outlays are essentially nonnegative.

Mathematically, the prospect that a sample data $x \in S$ belong to a class j, a value of a stochastic variable y, can be expressed as

$$P(y = j \mid x) = soft\max_j (b + Wx) = \frac{\exp(b_j + W_j x)}{\sum_i \exp(b_i + W_i x)} \quad \ldots (10)$$

The misclassification threshold values are introduced to turn the subsequent probabilities into class labels such that the misclassification outlays are lessened. By implementing the misclassification threshold value $1 - C_{i,j}$ on the obtained subsequent prospect $P(y = j \mid x)$, one can obtain the new prospect $P\xi$

$$P\xi(y = j \mid x) = P(y = j \mid x) \cdot (1 - C_{i,j}) \quad \ldots (11)$$

The hypothesized prediction $f(x)$ of the sample, x is the member of the maximum prospect among classes, can be obtained by using the following equation:

$$f(x) = \arg\max_j P\xi(y = j \mid x) \quad \ldots (12)$$

The proposed outlay-aware learning mechanism only concerns the output layer of a DBN. In this paper, we follow the same pre-training and enhancing procedures.

For imbalanced classification problems, the prior prospect distribution of different classes is essentially imbalanced or non-uniform. To reflect the class imbalance, there is a need to introduce the misclassification outlay at the output layer to reflect the imbalanced class distributions. In addition, traditional training algorithms generally assume uniform class distribution with equal misclassification outlays, i.e., $\forall i, j \in [1, 2, ..., K]$, if $i = j, C_{i,j} = 0$, if $i \neq j, C_{i,j} = 1$ which is not true in many real-world applications.

In many real-world applications, the misclassification outlays are essentially unknown, and they vary across various classes.

3.4. Evolutionary Outlay-Aware Deep Belief Network

EA is commonly used optimization algorithm. EA is inspired by the biological evolution process. The EA algorithm can be designed to optimize the misclassification outlays that are unknown in practice. In this paper, we propose an EOA-DBNC by incorporating outlay-aware function directly into its classification paradigm with the misclassification. The main idea of this outlay-aware learning technique is to assign class-dependent outlays. The procedure of training the proposed EOA-DBNC is summarized. First, a population of misclassification outlay is randomly initialized. We then train a DBN with the training data set. After applying misclassification outlays on the outputs of the DBN, we evaluate the training error based on the conductance of the corresponding outlay-aware hypothesized prediction. According to the evaluation conductance on training data set, proper misclassification outlays are selected to generate the population of next generation. In the next generation, mutation and crossover operators are employed to evolve a new population of misclassification outlays. Eventually, the best found misclassification outlays are obtained and applied to the output layer of DBN to form EOA-DBNC. During run time, we test the resulting EOA-DBNC with test data set to report the conductance. The practical steps of EOA-DBNC are summarized in Algorithm 1, and discussed next.

Training Process of EOA-DBNC

Pre-Training Phase:

1. Let S_t be the training dataset.
2. Train EOA-DBNC

Enhancing Phase:

1. Randomly initialize a population of misclassification outlays.
2. Generate a new population of misclassification outlays via mutation and crossover based on differential operator.
3. Multiply the corresponding misclassification outlays on training output and evaluate the error on training data.
4. According to evaluation conductance, select appropriate misclassification outlays and discard inappropriate ones in order to evolve next generation.
5. Continuously iterate between mutation and selection to reach the maximum number of generations.

3.4.1. Chromosome Encoding

Chromosome encoding is a significant step in EAs which aims at effectually demonstrating the imperative variables for improved conductance. In general, misclassification outlays in DBN are frequently indefinite. For getting apt outlays, in this work, each chromosome denotes the misclassification outlays for different classes, and the final evolved best chromosome is chosen as the misclassification outlays for EOA-DBNC. The chromosome encoding, here, directly encodes the misclassification outlays as values in the chromosome with numerical type and value range of [0, 1].

3.4.2. Population Initialization

The initial population is obtained via uniformly random sampling in feasible solution space for each variable within the specified range of the corresponding variable. The population holds the possible misclassification outlays and forms the unit of evolution. The evolution of the misclassification outlays is an iterative process with the population in each iteration called a generation.

3.4.3. Adaptive DE Operators

After initialization, adaptive differential evolution evolves the population with a sequence of three evolutionary operations, i.e., mutation, crossover, and selection, generation by generation. Mutation is carried out with DE mutation strategy to create the mutation discrete based on the current parent population as shown in Step 2.1 of Algorithm 1. After mutation, a binomial crossover operation is utilized to generate the final offspring as shown in Step 2.2 of Algorithm 1. In adaptive DE, each discrete has its associated crossover prospect instead of a fixed value. The selection operation selects the best one from the parent discrete and offspring discrete according to their corresponding fitness values as shown in Step 2.3 of Algorithm 1. Parameter adaptation is conducted at each generation. In this way, the control parameters are automatically updated to appropriate values without the need of prior parameter setting knowledge in DE. The crossover prospect of each discrete is generated independently based on a normal distribution with mean μC_r and standard deviation 0.1. Similarly, the mutation aspect of each discrete is generated independently based on a Cauchy distribution with location parameter μ_F and scale parameter 0.1. Both the mean μC_r and the location parameter μ_F are updated at the end of each generation as shown in Step 2.4 of **Algorithm 1.**

The chromosome is encoded with the misclassification outlays of different classes in numerical type. The evolution process mainly includes mutation, crossover, evaluation, and selection. The population is iteratively evolved via evolution process in each generation.

3.4.4. Evaluating the Fitness

Fitness is evaluated in order to select the suitable misclassification outlays. In this work, every discrete chromosome is introduced into discrete DBN as misclassification outlays. We generate suitable misclassification outlays for DBN using the training dataset. Mean value of training dataset is chosen as the objective function for the optimization.

3.4.5. Termination Condition

EAs are designed to evolve the population generation by generation and maintain the convergence as well as the diversity characteristics within the population. A maximum number of generations are set to be a termination condition of the algorithm. In this implementation, we consider the solutions converged when the best fitness value remains unchanged over the past 25 generations. The algorithm terminates either when it reaches the maximum number of generations or when it meets the convergence condition.

3.4.6. EOA-DBNC Creation

In due course, the optimization method split ends with the best discrete which is made use of misclassification outlays and accustom an EOA-DBNC. During the final stage of reaching the last generation, the best discrete is attained.

Algorithm - 2: Overall Working of EOA-DBNC

Input: X: Imbalanced records, Y: Class labels, N: Population size,
G: Maximum number of generations (i.e., stopping criterion), F: Mutation aspect,
Cr: Crossover prospect, D: Dimension of solution space, $range = (c_{\min}, c_{\max})$: Range of values for chromosome.
Set $\mu C_r = 0.5, \mu_F = 0.5, A = \phi, \beta = 0.5$

First Step - Initialization: Generate an initial population $\{c_0^1, ..., c_0^N\}$ via uniformly random sampling in solution space. The initial value of ithdiscrete is generated as $c_0^i = c_{\min} + rand(0,1) \cdot (c_{\max} - c_{\min})$. And evaluate each candidate solution $c_0^i (i = 1, ..., N)$ in the initial population via its corresponding trained $DBN(x \in S_t, y \mid c)$ to obtain a vector denoteing the fitness functions $F(c_0^i)$ which is G-mean of training data set in this paper.

Second Step - Evolution:

for g = 1, ... , G do

Set the set of all successful mutation aspect F_i at each generation $S_F = \phi;$

Set the set of all successful crossover probabilities Cr_i at each generation $S_{Cr} = \phi;$

for $i = 1, ..., N_p$ do

Generate $Cr_i = randn\ (\mu_{Cr}, 0.1)$,

$F_i = rand\, c_i\ (\mu_F, 0.1)$

Mutation: Randomly select two indices j and k from population, then generate a new candidate solution $c_g^{i'}$ from c_g^i, c_g^j and c_g^k by $c_g^{i'} = c_g^i + F_i \cdot \left(c_g^j - c_g^k\right)$ which is a DE operator.

Generate $i_{rand} = rand\,\mathrm{int}(1, D)$

Crossover:

if $i = i_{rand}\ or\ rand\,(0,1) < Cr_i$ then $u_g^i = c_g^{i'}$ g

else $u_g^i = c_g^i$

end if

Selection:

if $F\left(c_g^i\right) \geq F\left(u_g^i\right)$ then $c_{g+1}^i = c_g^i$

else $c_{g+1}^i = u_g^i, c_g^i \rightarrow A, Cr_i \rightarrow S_{Cr}, F_i \rightarrow S_F$

end if

Randomly remove solutions from A so that $|A| \leq N$

Parameter Adaptation:

$$\mu C_r = (1-\beta) \cdot \mu C_r + \beta \cdot mean\left(S_{CR}\right)$$

$$\mu_F = (1-\beta) \cdot \mu_F + \beta \cdot mean\left(S_F\right)$$

end for

end for

Third Step: EOA-DBNC Creation: Generate an EOA-DBNC with the best discrete c_{best} obtained from the training data set S_t as the misclassification cost.

Fourth Step: Run-time Evaluation: Evaluate EOA-DBNC on test dataset S_{test}.

IV. RESULTS AND DISCUSSIONS

Table 1. Dataset Name, No. of Instances and No. of Features

S.No	Dataset Name	No. of Instances	No. of Features
1	blood	748	5
2	bupa	345	7
3	car	1728	6
4	contraceptive	1473	9
5	credit	30000	24
6	diagnostic	569	32
7	ecoli	336	8
8	Ionosphere	351	34
9	mammography	961	6
10	monks - 1	432	7
11	monks - 2	432	7
12	monks - 3	432	7
13	Parkinson's	197	23
14	pima	768	8
15	prognostic	198	34
16	sonar	208	60
17	spect	267	22
18	Tic-Tac-Toe Endgame	958	9
19	vert	310	6
20	yeast	1484	8

Table 2. Predictive Accuracy of the Algorithms

Dataset	Algorithms				
	CBB	IGA - FLANN	ISVMC	APSO-CELMC	PSO – EOA-DBNC
blood	0.78	0.83	0.89	0.93	0.96
bupa	0.7	0.76	0.85	0.91	0.94
car	0.96	1	1	1	1
cont.	0.7	0.77	0.88	0.92	0.95
credit	0.86	0.89	0.94	0.97	0.99
diag.	0.95	0.99	1	1	1
ecoli	0.99	0.99	1	1	1
iono.	0.9	0.96	1	1	1
mamm.	0.82	0.85	0.94	0.97	0.99
mks-1	0.99	1	1	1	1
mks-2	0.68	1	1	1	1
mks-3	1	1	1	1	1
park	0.92	0.97	1	1	1
pima	0.75	0.81	0.93	0.96	0.99
prog.	0.77	0.84	0.95	0.95	0.98
sonar	0.8	0.89	0.95	0.96	0.99
spect	0.82	0.88	0.95	0.97	0.99
tic	0.94	1	1	1	1
vert	0.82	0.87	0.94	0.96	0.99
yeast	0.68	0.68	0.81	0.86	0.93

Table 3. Time Taken by the Algorithms for Classification (in milliseconds)

Dataset	Algorithms				
	CBB	IGA - FLANN	ISVMC	APSO-CELMC	PSO – EOA-DBNC
blood	2721	1032	649	772	715
bupa	3193	1948	1123	929	845
car	2795	1682	947	644	582
cont.	2696	1732	902	615	553
credit	19701	15782	8099	6580	6491
diag.	7790	3982	2184	1608	1547
ecoli	2686	1846	799	676	589
iono.	7663	3901	2001	1584	1519
mamm.	2672	1738	864	830	778
mks-1	2662	1936	895	692	612
mks-2	2757	1726	826	748	669
mks-3	2756	1639	904	705	648
park	4731	2393	1275	906	834
pima	2797	1392	780	535	477
prog.	2656	1888	947	674	609
sonar	8738	5189	2548	1723	1641
spect	5804	2291	1277	1066	981
tic	2807	1749	985	534	453
vert	2618	1638	849	671	585
yeast	2655	1843	909	576	488

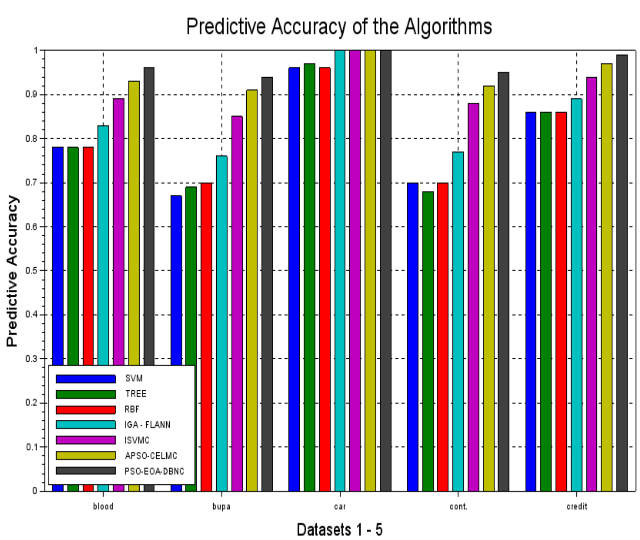


Fig.1. Predictive Accuracy Comparison for the datasets 1 to 5

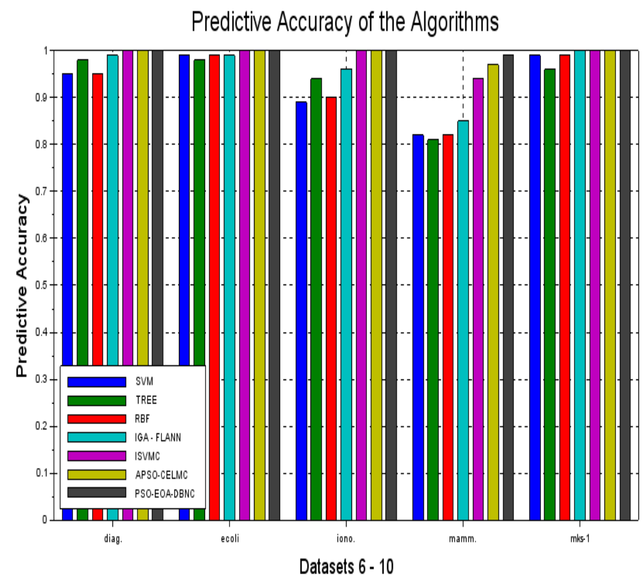


Fig.2. Predictive Accuracy Comparison for the datasets 6 to 10

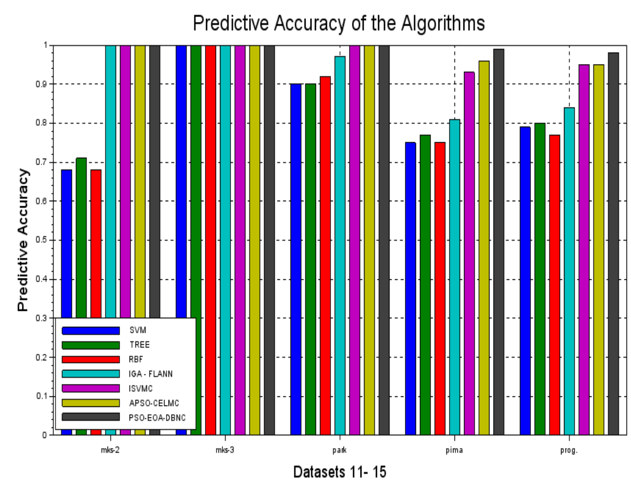


Fig.3. Predictive Accuracy Comparison for the datasets 11 to 15

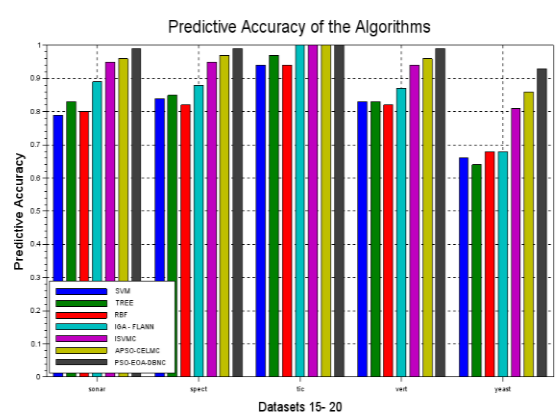


Fig.4. Predictive Accuracy Comparison for the datasets 16 to 20

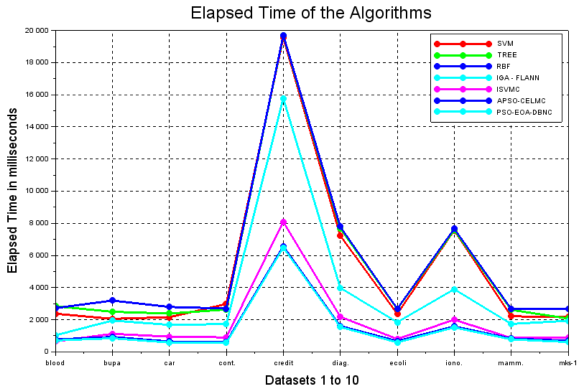
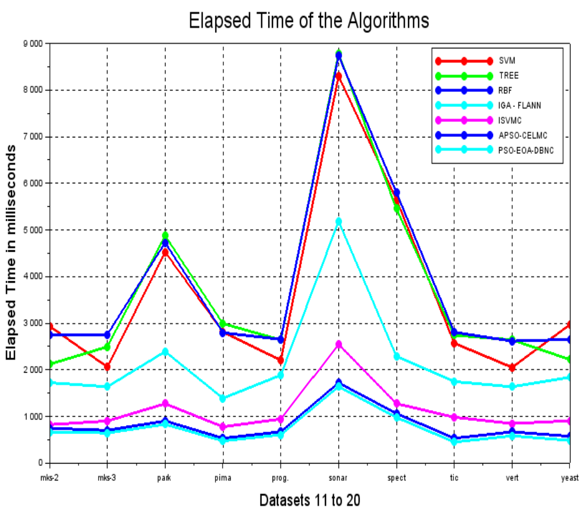


Fig.5. Time Taken for Classification by the Algorithms Comparison for the datasets 1 to 10



5. Time Taken for Classification by the Algorithms Comparison for the datasets 11 to 20

20 datasets are taken from the UCI machine learning repository namely blood, bupa, car, contraceptive, credit, diagnostic, ecoli, Ionosphere, mammography, monks – 1, monks – 2, monks – 3, parkinsons, pima, prognostic, sonar, spect, Tic-Tac-Toe Endgame, vert and yeast. The dataset details such as name of the dataset, number of instances and number of features are portrayed in the Table 1. The implementations are done using MATLAB tool. The system configuration is Core I3 processor with 8 GB RAM and 1 TB hard disk that runs on Microsoft Windows 8 operating system. Conductance metrics such as predictive accuracy and time taken for classification are obtained. For better visual depiction the results of the dataset is shown in 5 numbers. It is evident from the results that the proposed PSO-EOA-DBNC out conducts all the other algorithms in terms of prediction accuracy. Next, we compared the conductance of the proposed PSO-EOA-DBNC in terms of time taken for classification. From that results too, it is obvious that the proposed PSO-EOA-DBNC consumes less time than that of all the algorithms. The proposed PSO-EOA-DBNC algorithm is also compared with our previous works named IGA – FLANN [13], ISVMC [14] and APSO-CELMC [15]. It is significant PSO-EOA-DBNC that conducts better than that of IGA-FLANN, ISVMC and APSO-CELMC in terms of predictive accuracy and elapsed time for classification.

V. CONCLUSION

Two major tasks in knowledge discovery process are namely feature selection and classification. In this part of research work, particle swarm optimization is employed to carry out the task of feature selection. Classification is conducted by evolutionary outlay aware deep belief neural network. From the literature it is inferred that deep belief network conducts better in terms of elapsed time taken for classification task. For that reason, PSO based EOA-DBN is employed in this research work. 20 datasets are taken for conductance evaluation and from the obtained results it is evident that almost 97% accuracy is obtained and also the elapsed time taken for classification is reduced.

REFERENCES

[1] D. Polat, Z. Çataltepe, "Feature selection and classification on brain computer interface (BCI) data", in Proceedings of the 2012 20th Signal Processing and Communications Applications Conference (SIU), IEEE, 2012, pp. 1–4.
[2] G.P. Zhang, "Neural networks for classification: a survey", IEEE Trans. Syst. Man Cybern. Part C: Appl. Rev. 30 (4) (2000) 451–462.
[3] D.D. Lewis, "Naive (Bayes) at forty: the independence assumption in information retrieval, in Machine Learning": ECML-98, Springer, Berlin, Heidelberg, 1998, pp. 4–15.
[4] Kuan-Cheng Lin, Kai-Yuan Zhang, Yi-Hung Huang, Jason C Hung, Neil Yen, "Feature selection based on an improved cat swarm optimization algorithm for big data classification", J. Super computer. 72 (8) (2016) 3210–3221.
[5] Kuan-Cheng Lin, Yi-Hung Huang, Jason C. Hung, Yung-Tso Lin, "Feature selection and parameter optimization of support vector machines based on modified cat swarm optimization", Int. J. Distributed Sensational Network 2015 (2015).
[6] Kuan-Cheng Lin, Sih-Yang Chen, Jason C. Hung, "Feature selection and parameter optimization of support vector machines based on modified artificial fish swarm algorithms", Mathematical Probability Eng.(2015).
[7] C.A. Pena-Reyes, M. Sipper, "Evolutionary computation in medicine: an overview, Artificial Intelligence", Med. 19 (1) (2000) 1–23.
[8] S.H. Cha, C. Tappert, "A genetic algorithm for constructing compact binary decision trees", J. Pattern Recognition. Res. 4 (1) (2009) 1–13.
[9] J. Kennedy, "Particle swarms optimization, in Encyclopaedia of Machine Learning", Springer, US, 2010, pp. 760–766.
[10] P.P. Brahma, D. Wu, Y. She, "Why Deep Learning Works: A Manifold Disentanglement Perspective", IEEE Transactions on Neural Networks & Learning Systems, 2016, 27(10):1997-2008.
[11] D. Li, S. Y. Dong, "Deep learning: methods and applications, Foundations & Trends in Information Retrieval", 2014, 7(3):197-387.

[12] R. Salakhutdinov, G. Hinton," An efficient learning procedure for deep Boltzmann machines, Neural Computation", 2012, 24(8):1967.
[13] M.Praveena, Dr.V.Jaiganesh, "Improved Genetic Algorithm Based Feature Selection Strategy Based Five Layered Artificial Neural Network Classifier (IGA – FLANN)", International Journal of Engineering and Techniques - Volume 3 Issue 5, Sep - Oct 2017, 199-213.
[14] M.Praveena, Dr.V.Jaiganesh, "Routine Correspondence Method with Grey Wolf Optimization based Imperforate Support Vector Machine Classifier (ISVMC) for High Dimensional Datasets", Journal of Advanced Research in Dynamical & Control Systems, Vol. 11, 01-Special Issue, 2019, 652-660.
[15] M.Praveena, Dr.V.Jaiganesh, "Adaptive Particle Swarm Optimization based Credentialed Extreme Learning Machine Classifier (APSO-CELMC) for High Dimensional Datasets", International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8, Issue-10S, August 2019.
[16] M.Praveena, Dr.V.Jaiganesh, "A Literature Review on Supervised Machine Learning Algorithms and Boosting Process", International Journal of Computer Applications (0975 – 8887) Volume 169 – No.8, July 2017.

Authors Profile

Mrs.M.Praveena pursed Bachelor of Science from Avinashilingam Deemed University, Coimbatore in 2003 and Master of Computer Applications from Bharathiar University in year 2006. She Pursue Master of Philosophy in Computer Science from Periyar University, Salem in 2007. She is currently pursuing Ph.D in Dr.NGP Arts and Science College, Coimbatore and currently working as Assistant Professor in Department of Computer Science, Dr.SNS Rajalakshmi College of Arts and Science, Coimbatore, since 2006.She is a life member of the The Indian Science Congress Association, Kolkata since 2010.She has published more than 20 research papers in reputed international journals including Journal of Advanced Research in Dynamical & Control Systems and presented paper in conferences including IEEE. Her main research work focuses on Data Mining and bio inspired algorithms. She has 13 years of teaching experience and 4 years of Research Experience. She guided 3 M.Phil Research Scholars and completed 5 online NPTEL courses.

Dr. V. Jaiganesh pursed Bachelor of Science and Master of Computer Applications from Salem Sowdeswari College, Salem in 1998 and 2001. He Pursued Master of Philosophy in Computer Science from Periyar University, Salem in 2007 and Ph.D from Manonmaniam Sundaranar University, Tirunelveli in 2015. He is currently working as Professor in Department of Computer Science, Dr.NGP Arts and Science College, Coimbatore, since 2007. He is an Editorial Board Member in various reputed journal like International Arab Journal of Information Technology, since 2015, Editorial Board Member in International Journal of Research in Engineering and Technology, since 2015 and Editorial Board Member in International Journal of Computer Technology and Applications, since 2015. He has published more than 20 research papers in reputed international journals including IEEE Digital Library and presented paper in various conferences including IEEE International conference on ICICES, SA Engineering College, Chennai. He got Recognition and Appreciation for outstanding service in the Technical Committee of IAJIT as one of the Active Reviewers. His main research work focuses on Wireless Sensor Network and Data Mining. He has 19 years of teaching experience and 13 years of Research Experience. He guided 5 M.Phil Research Scholars and guiding 7 Ph.D Research Scholars.

