2020 International Conference on Computational Performance Evaluation (ComPE)
North-Eastern Hill University, Shillong, Meghalaya, India. Jul 2-4, 2020

Feature Selection Using Metaheuristic Algorithms: Concept, Applications and Population Based Comparison

Rahul Hans
Department of Computer Science and Engineering,
Guru Nanak Dev University, Regional Campus, Gurdaspur,
Punjab, India
rahulhans@gmail.com

Harjot Kaur
Department of Computer Science and Engineering,
Guru Nanak Dev University, Regional Campus, Gurdaspur,
Punjab, India
harjotkaursohal@rediffmail.com

***Abstract*— Technologies like machine learning in the current times, have emerged as capable domains of research in Computer Science. In Machine learning, the system is trained on the basis of the available dataset; the dataset may contain many redundant and irrelevant features which may require more memory for storage and also increases the cost of computation. Selection of best features enhances the accuracy of data classification along with working on smallest amount of features is considered as an optimization problem. Metaheuristic algorithms in current times have been used far and wide unravel various optimization problems. In this context, this study aims to discuss the solution of feature selection problem using metaheuristic algorithms and presents a population based comparison of four metaheuristic algorithms for extracting smallest feature subset with utmost accuracy.**

***Keywords*— Machine Learning, Feature Selection, Optimization, Metaheuristic Algorithms, K-NN.**

I. INTRODUCTION

In this era of technology, an outsized sum of data is spawned and is brought together in the form of immense sum features to form a deposit/set of data. Features are basically worth, a process is having [1]. These sets of data contains enormous features which are irrelevant and may not be appropriate for the problem under consideration [2] that require additional space for the storage and lead to elevate computational overheads. As a consequence, obtaining higher accuracy in the classification with small amount of selected features to decrease computational/storage overheads, feature selection is key task. In modern era a lot of effort has been made by researchers to trim down the number of chosen features while capitalize on the accuracy of classification of the data. Diverse metaheuristic algorithms have been applied to accomplish this purpose. This paper aims to summarize some existing research done in the field of choosing least number of features by means of metaheuristic algorithms. Section 2 presents briefly the Inspiration of the reserach, where as Section 3 throws light on the Binary Optimization and feature selection. Furthermore Section 4 presents some of the applications where feature selection using metaheurstic algorithms has been applied successfully. Section 5 presents materials and methods and section 6 presents the results for two different population sizes and lastly some future directions with conclusions is presented in section 7.

II. INSPIRATION

Selection of optimal number of features is a tedious job, due to the complexity of the search process for finding all feasible solutions in a good enough time. This process can be modeled as a problem of optimization [3]. Metaheuristic algorithms in current era have proved their potential in solving various optimization problems. Basically, these algorithms need to uphold stability in two of the following significant mechanisms: The mechanism of exploring the search domain of solutions along with the mechanism for exploiting the superlative solution brought into being in the preceding phase or the former mechanism produces the dissimilar results by traveling around in the search domain of solutions on large-scale, while the later one ponders over the hunt for the solution in the confined domain by bearing in mind the fact; existing premium result might be present in this sub domain.

III. BINARY OPTIMIZATION TASK AND FEATURE SELECTION

The problem under consideration can be mapped as a binary translation of the continuous optimization problem where the solutions of the problem(initial population/search agents) can be represented in the form of vectors of 0's and 1's as shown in figure 1, where SA represents the search agent and feature set is represented by $(F_1, F_2, F_3 \ldots\ldots F_n)$, for a particular search agent if the value of feature F_i is 1 that means that solution selects that feature, while the value 0 means the converse. This lays the ground to model the problem under consideration as a binary edition of continuous optimization problem in which we select the features from a bigger dataset or we don't select the features for the evaluation for the fitness which can be modeled as 1's for all the selected features and 0's for all the unselected features. The continuous valued search space can be altered to

978-1-7281-6644-5/20/$31.00 ©2020 IEEE

binary search space by transforming their variables containing continuous values to binary values 0s and 1s. For a binary problem the search space is represented by sequence of 0s and 1s strings, to symbolize the domain of search of a binary task a hypercube structure is modeled in the literature where tuning the individual bit of the candidate solution (search agent/ant/ant lion/universe/chromosome), could shift it to close or far away

corner of the modeled hypercube structure. [4, 5].

	F_1	F_2	F_3	F_4	F_5	F_6	F_7	.	.	F_n
SA_1:	1	0	1	1	0	0	1	.	.	1
SA_2:	0	1	0	1	0	0	1	.	.	1
SA_3:	0	1	1	1	1	0	1	.	.	1
SA_4:	1	0	0	1	0	0	1	.	.	1
SA_5:	1	0	1	1	0	1	1	.	.	1
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
SA_k:	1	0	0	1	0	0	1	.	.	1

Fig. 1. Binary Feature set for Population size k with n features

The modeling of considered continuous algorithm for optimization to its contantrary binary version becomes a compulsory prerequisite.

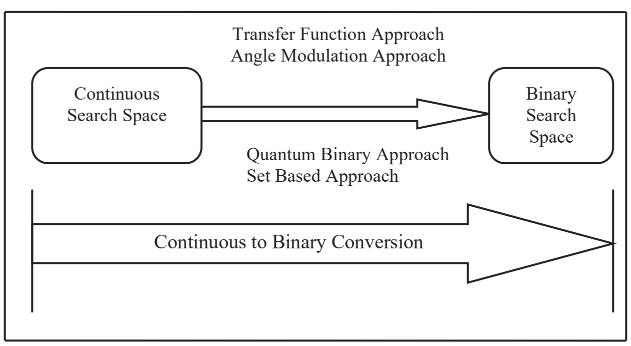

Fig. 2. Mapping of Continuous Search Space to Binary Search Space

It can be acknowledged from the literature that the techniques for the binary conversion are divided among two distinct groups, the first one presents the general methods for binary conversion. In this group the alteration is accomplished without varying algorithms specific operators[10]. Two approaches have been considered for this viz. approach based on transfer functions and an approach based on the concept of Angle modulation . The second group corresponds to those methods that are developed especially for a metaheuristic, within them techniques like set based approach along with the quantum binary approach is there. As per the study conducted in [4] transfer functions have noteworthy affect on intensification and diversification processes as discussed above. So a careful selection of transfer function is needed for enhancement of the performance.

IV. Applications

In [11] author considers the problem of **Image steganalysis** that is to find the secret information from images or from any multimedia file. For this an Adaptive inertia weight dependent concept is incorporated with Particle Swarm Optimization which helps to make the best combination of features and also remove the unwanted ones is presented. The author improved this algorithm by adaptively adjusting the inertia weights of PSO. In [12] for selection of relevant features in **obstructive sleep apnea**, authors proposed the combination of PSO with 1-NN method. For performance validation of the presented technique author also make use of the 8 life sciences datasets form UCI

repository. In [13] for applying the feature selection process the phases for **categorization of cancer** is presented in [14], in which the Feature Selection based on Correlation- along with the author combined PSO algorithm with KNN classifier which is tested on 11 datasets for gene expression. A model based on two improved version of Binary PSO is proposed. Eleven datasets of different cancer types have been used for evaluation. Cat swarm optimization is improved in [15] to present an improved version of cat swarm optimization for selecting features for **text classification in big data**. Authors proposed two variants of cat swarm optimization by amending the mode of seeking in the first step and replaced the original technique used to change the location of the cats. Rapid convergence was seen by modifying the algorithm as discussed above. In [16] authors proposed **Facial Expression Recognition** technique by using differential evolution algorithm, three different datasets related to the Facial Expression Recognition system is used for evaluation. In [17] author proposed the hybrid version of supervised particle swarm optimization. The author combines the PSO and rough set theory for selecting features to **diagnosis of diseases** or medical datasets. Now days Computer-aided diagnosis (CAD) is a indispensible part of medical domain it helps in pre determine the any kind of disease by just looking at the images. In this paper [19] for **diagnosis of bronchitis** author proposed hybrid version of ACO. Author used the ant colony optimization algorithm along with the technique based on the cosine similarity known as with tandem run recruitment. SVM is used to check the strength of feasible solutions generated so far. This hybrid version is tested on bronchitis dataset. The result in this work shows that proposed algorithm provides better classification accuracy. In [7] author proposed the hybrid version of Whale Optimization Algorithm with local search mechanism based on Simulated Annealing. They both are combined in a way that after every iteration when WOA found some regions by means of exploring then there come a role of Simulated Annealing.

V. Materials And Methods

978-1-7281-6644-5/20/$31.00 ©2020 IEEE

The problem of selection of features from a bigger set is understood to be as an optimization problem with two different objectives also known as multi-objective optimization [8]. For the successful execution of the algorithm, resultant solution should consider minimum number of features while improving the accuracy [9]. k-NN classifier is employed to worth the fitness of the candidate solutions shown in (1).

$$Fitness = X_1 \left|\frac{FS}{N}\right| + X_2 * \Omega(D) \quad \ldots\ldots (1)$$

No.	Dataset	Instances	Attributes
D1	Zoo	101	16
D2	Statlog Credit	1000	20
D3	Heart	294	13
D4	Lung cancer	32	56

Where $\Omega(D)$ symbolize the rate of error during the process of classification computed from k-NN classifier. |FS| symbolizes selected feature subset and exact feature set size is | N |, X_1 and X_2 signifies the prominence, The value of $X_1 \in [0, 1]$ and $X_2 = (1 - X_1)$ adopted from [7, 6, 20].

A. Parameter Settings and Dataset

The experiments have been performed on CPU with 3.00 GB RAM and Core 2 Duo 2.00 GHz. To carry out the experiments, four datasets were attained from online sources [64, 65] as summarized in Table 2.

Parameter	Value
Total Runs of Algorithm	10
Total Iterations taken	100
Initial population Size	10 and 20
Lower bound	0
Upper bound	1
Dimensions	Total features in dataset

TABLE 1: Parameter Settings

TABLE 2: Dataset Description

B. Performance Metrices

- Mean Accuracy(*Mean_Accu*)- It is classification accuracy averaged on all the runs.
- Mean fitness(*Mean_Fitn*) – It is the fitness values averaged on all the runs.
- Worst fitness(*Wst_Fitn*) –It is the maximum among all fitness values in all runs.
- Best fitness(*Bst_Fitn*) - It is the minimum among all fitness values in all runs.
- Mean Features Selected(*Mean_Ftr_Sel)* – It is elected features averaged on all the runs.
- Mean standard Deviation(*Mean_Std_Dev*)- It is the standard deviation values averaged on all the runs.
- Average Time Taken(*Mean_Time/Run*)- It is the average of computational time in all the runs.

		Mean_ Accu	***Mean_ Fitn***	***Wst_ Fitn***	***Bst_ Fitn***
GA	Zoo	0.927	0.077	0.158	0.024
	Statlog Credit	**0.733**	0.269	0.288	0.245
	Heart	**0.822**	**0.179**	0.254	0.112
	Lung cancer	**0.950**	**0.052**	0.127	0.003
MVO	Zoo	**0.945**	**0.061**	0.177	0.026
	Statlog Credit	0.717	0.287	0.302	0.265
	Heart	0.812	0.190	0.228	0.134
	Lung cancer	0.863	0.142	0.314	0.068
GWO	Zoo	0.871	0.135	0.236	0.043
	Statlog Credit	0.711	0.291	0.308	0.258
	Heart	0.809	0.193	0.248	0.138
	Lung cancer	0.856	0.147	0.315	0.066
BBO	Zoo	0.904	0.099	0.158	0.043
	Statlog Credit	0.725	0.277	0.291	0.253
	Heart	0.814	**0.187**	0.213	0.164
	Lung cancer	0.881	0.120	0.373	0.002

VI. Analysis Of Results And Discussions

In this segment all the results acquired by comparing the performance of four different metaheuristic algorithms are reported viz Genetic algorithm(GA), Grey Wolf optimization

TABLE 3: Analysis of the algorithms for Population Size 10 for Accuracy and Fitness values

		Mean _Std_Dev	***Mean_Ftr_Sel***	***Mean Time/ Run***
GA	Zoo	0.050	**6.8**	18.39
	Statlog Credit	0.014	8.8	43.73
	Heart	0.017	4.2	18.36
	Lung cancer	0.072	14.3	27.70
MVO	Zoo	0.041	11.1	23.13
	Statlog Credit	0.013	12.8	38.06
	Heart	0.022	7.7	18.16
	Lung cancer	0.039	33.4	17.27
GWO	Zoo	0.037	10.3	20.11
	Statlog Credit	0.012	11.3	45.68
	Heart	0.033	5.2	22.81
	Lung cancer	0.076	24	24.16
BBO	Zoo	0.040	7.1	23.01
	Statlog Credit	0.017	**7.6**	45.35
	Heart	0.020	**3.9**	24.52
	Lung cancer	0.045	**9.3**	22.59

TABLE 4: Analysis of the algorithms for Population Size 10 for Standard deviation, Features Selected and Time taken

		Mean _Std_Dev	***Mean_Ftr_Sel***	***Mean Time/ Run***
GA	Zoo	0.038	8	10.01
	Statlog Credit	0.014	8.4	14.31
	Heart	0.039	4.5	10.23
	Lung cancer	0.057	16.1	9.71
MVO	Zoo	0.045	9.8	9.73
	Statlog Credit	0.011	13.8	19.75
	Heart	0.030	5.9	10.86
	Lung cancer	0.081	33.6	9.43
GWO	Zoo	0.058	10.3	10.24
	Statlog Credit	0.013	9.9	19.62
	Heart	0.032	5.4	11.81
	Lung cancer	0.088	25.8	9.78
BBO	Zoo	0.038	**7**	10.24
	Statlog Credit	0.012	**8.2**	13.99
	Heart	0.018	**4**	10.24
	Lung cancer	0.125	**14**	9.05

978-1-7281-6644-5/20/$31.00 ©2020 IEEE

algorithm(GWO), Multi-verse optimization algorithm (MVO) and Biogeography based optimization algorithm(BBO) .

TABLE 5: Analysis of the algorithms for Population Size 20 for Accuracy and Fitness values

		Mean_ Accu	*Mean_ Fitn*	*Wst_ Fitn*	*Bst_ Fitn*
GA	Zoo	**0.939**	**0.065**	0.160	0.004
	Statlog Credit	0.720	0.282	0.303	0.263
	Heart	0.830	0.172	0.199	0.144
	Lung cancer	0.919	0.083	0.250	0.002
MVO	Zoo	**0.939**	0.067	0.140	0.004
	Statlog Credit	0.721	0.282	0.299	0.258
	Heart	0.834	0.170	0.200	0.131
	Lung cancer	0.888	0.117	0.192	0.067
GWO	Zoo	0.919	0.087	0.162	0.043
	Statlog Credit	0.717	0.286	0.305	0.258
	Heart	0.805	0.197	0.242	0.150
	Lung cancer	0.863	0.140	0.251	0.005
BBO	Zoo	0.937	0.067	0.160	0.025
	Statlog Credit	**0.746**	**0.256**	0.283	0.233
	Heart	**0.837**	**0.164**	0.192	0.124
	Lung cancer	**0.944**	**0.057**	0.125	0.001

and it is clearly understood from the results that with the increase in the population size the performance increases but the computation time increases. The table 3 and table 4 depicts the performance of the algorithms considered for the population size 10 and table 5 and table 6 depicts the performance with initial population sizes 20 based on the above metrics. With the increase in the size of initial population the mean accuracy of classification averaged on all the datasets increases and features selected summed on all the datasets is reduced.

TABLE 6: Analysis of the algorithms for Population Size 20 for Standard deviation, Features Selected and Time taken

V

This research presents brief overview of the concept of one of the optimization problem viz. feature selection. The proposed study summarizes some of the recent applications in which the authors proposed to solve this problem by exploring the use of some of the recent metaheuristic algorithms. Selection of minimum features can be considered as a multiobjective optimization problem having tradeoffs among two conflicting objectives of minimum features and maximum accuracy. The proposed research analyses the performance of the metaheuristic algorithms for two different population sizes the results clearly indicate that the efficiency of the algorithms increases with the increase in the population size. In the future, work can be done to implement optimization algorithms to unravel feature selection for specific application. Also work can be done to diminish the time of computation.

REFERENCES

[1]. G. Chandrashekar and F. Sahin “A survey on feature selection methods, “ *Computers & Electrical Engineering*, *40*(1), 16-28, 2014.

[2]. E. Hancer, B. Xue, D. Karaboga and M. Zhang “A binary ABC algorithm based on advanced similarity scheme for feature selection,” *Applied Soft Computing*, *36*, 334-348, 2015.

[3].X.S. Yang, S. Deb and S. Fong “Metaheuristic algorithms: optimal balance of intensification and diversification,” Applied Mathematics & Information Sciences, 8(3), 977, 2014.

[4].S. Saremi, S. Mirjalili, and A. Lewis “How important is a transfer function in discrete heuristic algorithms,” *Neural Computing and Applications*, 26(3), 625-640, 2015.

[5]. S. Mirjalili, and A. Lewis “S-shaped versus V-shaped transfer functions for binary particle swarm optimization,” *Swarm and Evolutionary Computation*, 9, 1-14, 2013.

[6]. E. Emary, H. M. Zawbaa and A. E. Hassanien “Binary ant lion approaches for feature selection” Neurocomputing, 213, 54-65, 2016.

[7].M.M. Mafarja and S. Mirjalili “Hybrid Whale Optimization Algorithm with simulated annealing for feature selection”. Neurocomputing, 260, 302-312,2017.

[8]. B. Xue, M. Zhangand W.N. Browne “Particle swarm optimization for feature selection in classification: A multi-objective approach” *IEEE transactions on cybernetics*, *43*(6), 1656-1671,2013.

[9]. M. A. Hall “Correlation-based feature selection for machine learning”,1999.

[10]. B. Crawford, R. Soto, G. Astorga, J. Garcia, C. Castro, F. Paredes “Putting continuous metaheuristics to work in binary search spaces,” Complexity, 2017.

[11]. A. Adeli and A. Broumandnia “ Image steganalysis using improved particle swarm optimization based feature selection,” *Applied Intelligence*, *48*(6), 1609-1622,2018.

[12]. L.F. Chen, C.T. Su, K.H. Chen, and P.C. Wang “Particle swarm optimization for feature selection with application in obstructive sleep apnea diagnosis,” *Neural Computing and Applications*, *21*(8), 2087-2096,2012.

[13]. L.Y. Chuang, H.W. Chang, C.J. Tu, and C.H. Yang “Improved binary PSO for feature selection using gene expression data,” *Computational Biology and Chemistry*, *32*(1), 29-38,2008.

[14]. I. Jain, V.K. Jain, and R. Jain “Correlation feature selection based improved-binary particle swarm optimization for gene selection and cancer

978-1-7281-6644-5/20/$31.00 ©2020 IEEE

classification," *Applied Soft Computing*, *62*, 203-215, 2018.

[15]. K.C. Lin, K.Y. Zhang, Y.H. Huang, J.C. Hung and N. Yen "Feature selection based on an improved cat swarm optimization algorithm for big data classification," *The Journal of Supercomputing*, *72*(8), 3210-3221,2016.

[16]. U. Mlakar, I. Fister, J. Brest, and B. Potočnik "Multi-objective differential evolution for feature selection in facial expression recognition systems," *Expert Systems with Applications*, *89*, 129-137,2017.

[17]. H.H. Inbarani, A.T. Azar, and G. Jothi, "Supervised hybrid feature selection based on PSO and rough sets for medical diagnosis," *Computer methods and programs in biomedicine*, *113*(1), 175-185. 2014.

[18]. L. Shang, Z. Zhou and X. Liu "Particle swarm optimization-based feature selection in sentiment classification," *Soft Computing*, *20*(10), 3821-3834,2016.

[19]. J.D. Sweetlin, H.K. Nehemiah, and A. Kannan "Feature selection using ant colony optimization with tandem-run recruitment to diagnose bronchitis from CT scan images," *Computer methods and programs in biomedicine*, *145*, 115-125,2017

978-1-7281-6644-5/20/$31.00 ©2020 IEEE

978-1-7281-6644-5/20/$31.00 ©2020 IEEE