Bio-Inspired Feature Selection: An Improved Binary Particle Swarm Optimization Approach

Background: Feature selection is an effective approach to reduce the number of features of data, which enhances the performance of classification in machine learning. The goal of feature selection in this work is to reduce the number of selected features while maximizing the classification accuracy, which can be regarded as a multi-objective optimization problem**.**

Problem: In this paper, we formulate a joint feature selection problem to reduce the number of the selected features while enhancing the accuracy.

Method: An improved binary particle swarm optimization (IBPSO) algorithm is proposed to solve the formulated problem.

Experimental Setting: In the feature selection tests, the genetic algorithm (GA), binary firefly algorithm (BFA), binary cuckoo search (BCS), BPSO, and binary bat algorithm (BBA) are introduced as the comparison algorithms. Moreover, the key parameter setups of these comparison algorithms as well as the proposed IBPSO are listed in Table 3. In addition, the maximum number of iterations for each algorithm is set as 200, the population size (number of searching agents) is 20, and the dimension of solution is equal to the feature number of each dataset. Note that the performance of a metaheuristic's algorithm is directly affected by the population size and the number of iterations. Specifically, if the algorithm is with large population size, it may achieve better optimization performance than the algorithm with small population size. Moreover, if an algorithm has more numbers of iterations, then it may obtain better results than the algorithm with less numbers of iterations. Thus, we use the same population size and the number of iterations for each algorithm to make a fair comparison between different algorithms.

Algorithm	Values of key parameters
GA	Mutation rate $= 0.05$, Crossover rate $= 0.5$
BFA	Random coefficient $= 0.1$, Absorption coefficient $= 0.3$
BCS	Discovery probability $= 0.25$, $\alpha = 1$
BPSO	$c_1 = 1.5$, $c_2 = 2.0$
BBA	$A = 0.25$, $Q_{max} = 2$, $Q_{min} = 0$
IBPSO	$c_1 = 1.75$, $c_2 = 2.0$, $\alpha = 1$, $\omega_{max} = 2$, $\omega_{min} = 0$

TABLE 3 Parameter Setups of Different Algorithms

Each algorithm is independently run for 30 times to solve the feature selection problems of these selected datasets, and the numerical statistics results will be presented. Moreover, in each test, we use 80 % of the instances for training, and the rest ones are used for testing, which is a common way adopted by several previous works.

The computer used for the tests is with an Intel(R) Xeon(R) E5-2630 v4 CPU and the RAM is 32 GB. Moreover, the abovementioned algorithms for feature selections are implemented by Python.

Experimental Result:

Performance Evaluations of Different Algorithms: The results of the selected 16 datasets are shown in two separated tables. It can be seen from the tables that the proposed IBPSO algorithm achieves the best average fitness function values on 12 datasets, which means it has better performance than other comparison algorithms.

Feature Selection Accuracies: Similarly, the numerical statistics results of different algorithms for each dataset are presented in these tables. As can be seen, IBPSO algorithm achieves the best average accuracies of feature selection results on 10 datasets and the best accuracy results on 13 datasets. Thus, IBPSO algorithm has the best performance in terms of feature selection accuracies on these selected datasets compared to other algorithms.

Dataset	Algorithm	Best	Worst	SD	Mean	CPU Time
Breastcancer	GA	**0.0272**	**0.0272**	**0.0000**	**0.0272**	69.9977
	BFA	**0.0272**	0.0309	0.0013	0.0288	138.5477
	BCS	**0.0272**	**0.0272**	**0.0000**	**0.0272**	**20.8977**
	BPSO	**0.0272**	**0.0272**	**0.0000**	**0.0272**	136.9305
	BBA	**0.0272**	**0.0272**	**0.0000**	**0.0272**	142.7571
	IBPSO	**0.0272**	**0.0272**	**0.0000**	**0.0272**	146.9765
BreastEW	GA	**0.0389**	0.0416	0.0008	0.0396	**83.9037**
	BFA	0.0399	0.0412	0.0004	0.0405	168.6450
	BCS	0.0433	0.0470	0.0010	0.0448	179.9466
	BPSO	0.0399	0.0409	0.0003	0.0404	171.6200
	BBA	0.0392	0.0399	0.0002	0.0396	167.6903
	IBPSO	**0.0389**	**0.0392**	**0.0002**	**0.0390**	168.3109
Exactly	GA	**0.0046**	**0.0046**	0.0000	**0.0046**	**100.7601**
	BFA	**0.0046**	0.0173	0.0038	0.0059	188.3626
	BCS	**0.0046**	0.0123	0.0028	0.0086	115.8868
	BPSO	**0.0046**	**0.0046**	**0.0000**	**0.0046**	198.0611
	BBA	**0.0046**	**0.0046**	**0.0000**	**0.0046**	199.2861
	IBPSO	**0.0046**	**0.0046**	**0.0000**	**0.0046**	197.6734
Exactly2	GA	**0.2097**	0.2260	0.0057	0.2132	**97.3038**
	BFA	**0.2097**	**0.2097**	**0.0000**	**0.2097**	178.1391
	BCS	**0.2097**	0.2309	0.0081	0.2173	117.0541
	BPSO	**0.2097**	**0.2097**	**0.0000**	**0.2097**	186.8110
	BBA	**0.2097**	**0.2097**	**0.0000**	**0.2097**	177.7468
	IBPSO	**0.2097**	**0.2097**	**0.0000**	**0.2097**	184.6734
HeartEW	GA	**0.1490**	0.1648	0.0061	0.1529	**57.3660**
	BFA	**0.1490**	0.1607	0.0045	0.1525	113.1633
	BCS	**0.1490**	0.1638	0.0051	0.1572	62.2022
	BPSO	**0.1490**	**0.1490**	**0.0000**	**0.1490**	114.3146
	BBA	**0.1490**	**0.1490**	**0.0000**	**0.1490**	114.6943
	IBPSO	**0.1490**	**0.1490**	**0.0000**	**0.1490**	117.7609
Lymphography	GA	**0.5385**	0.5836	0.0136	0.5602	**54.3051**
	BFA	0.5511	**0.5660**	**0.0046**	0.5624	108.4117
	BCS	0.5583	0.5798	0.0063	0.5703	103.0773
	BPSO	0.5583	0.5764	0.0063	0.5614	108.7586
	BBA	**0.5385**	0.5704	0.0100	0.5558	108.5527
	IBPSO	**0.5385**	0.5704	0.0094	**0.5551**	112.4991
SonarEW	GA	0.0747	0.1075	0.0089	0.0938	**69.3591**
	BFA	0.0750	0.1038	0.0076	0.0947	138.3002
	BCS	0.1092	0.1198	**0.0041**	0.1151	145.1466
	BPSO	0.0946	0.1035	0.0029	0.0994	142.5023
	BBA	0.0754	0.0993	0.0066	0.0849	141.3256
	IBPSO	**0.0547**	**0.0929**	0.0113	**0.0749**	148.5394
SpectEW	GA	0.2534	0.2846	0.0100	0.2644	**57.9660**
	BFA	0.2603	0.2763	0.0055	0.2700	114.3705
	BCS	0.2589	0.2850	0.0080	0.2751	116.4804
	BPSO	0.2585	0.2731	**0.0048**	0.2661	116.6135
	BBA	0.2488	**0.2662**	0.0053	**0.2575**	116.1148
	IBPSO	**0.2507**	0.2777	0.0079	0.2623	119.6393

TABLE 4 Fitness Values Obtained by Different Algorithms (Datasets 1 to 8)

Dataset	Algorithm	Best	Worst	SD	Mean	CPU Time
CongressEW	GA	**0.0224**	0.0279	0.0017	0.0256	**75.2692**
	BFA	**0.0224**	0.0285	0.0016	0.0252	146.5245
	BCS	0.0269	0.0320	0.0016	0.0289	132.6274
	BPSO	**0.0224**	0.0256	**0.0013**	**0.0237**	148.4214
	BBA	**0.0224**	0.0281	0.0023	0.0249	148.0824
	IBPSO	**0.0224**	**0.0275**	0.0020	0.0243	151.8781
IonosphereEW	GA	0.0846	0.1141	0.0105	0.0997	**78.1663**
	BFA	0.0901	0.1154	0.0070	0.1052	154.8571
	BCS	0.1023	0.1384	0.0095	0.1270	167.1765
	BPSO	0.0961	0.1132	**0.0059**	0.1050	157.9756
	BBA	0.0791	**0.1066**	0.0088	0.0943	153.9433
	IBPSO	**0.0705**	0.1147	0.0140	**0.0905**	159.0922
KrvskpEW	GA	0.0238	0.0359	0.0034	0.0295	**982.3659**
	BFA	0.0338	0.0454	0.0032	0.0389	1617.9190
	BCS	0.0324	0.0403	**0.0023**	0.0379	2082.1176
	BPSO	0.0276	0.0371	0.0026	0.0322	1915.5977
	BBA	0.0259	0.0353	0.0029	0.0286	1982.8664
	IBPSO	**0.0210**	**0.0276**	0.0024	**0.0241**	2116.3775
Tic-tac-toe	GA	0.1564	0.1564	0.0000	0.1564	89.3500
	BFA	0.1564	0.1564	0.0000	0.1564	188.8746
	BCS	0.1564	0.1564	0.0000	0.1564	**13.8251**
	BPSO	0.1564	0.1564	0.0000	0.1564	177.2706
	BBA	0.1564	0.1564	0.0000	0.1564	196.1019
	IBPSO	0.1564	0.1564	0.0000	0.1564	204.5367
Vote	GA	**0.0493**	0.0630	0.0033	0.0558	**58.6129**
	BFA	**0.0493**	0.0590	0.0028	0.0535	115.3932
	BCS	0.0551	0.0611	0.0024	0.0579	102.7837
	BPSO	**0.0493**	0.0539	0.0017	0.0503	117.6988
	BBA	**0.0493**	**0.0520**	**0.0008**	**0.0496**	116.0473
	IBPSO	**0.0493**	0.0539	0.0016	0.0509	118.0201
WaveformEW	GA	0.1577	0.1761	0.0054	0.1677	**3739.2315**
	BFA	0.1778	0.1834	**0.0019**	0.1810	6207.9586
	BCS	0.1715	0.1809	0.0027	0.1764	9771.4048
	BPSO	0.1644	0.1769	0.0037	0.1721	7674.0827
	BBA	0.1625	0.1733	0.0035	0.1669	7524.7088
	IBPSO	**0.1573**	**0.1698**	0.0033	**0.1621**	7356.3544
Zoo	GA	**0.0584**	0.0944	0.0107	0.0702	**53.8582**
	BFA	**0.0584**	0.0757	0.0060	0.0639	108.6919
	BCS	**0.0584**	0.0764	0.0055	0.0668	89.2766
	BPSO	**0.0584**	0.0674	0.0027	**0.0593**	105.4454
	BBA	**0.0584**	0.0680	0.0028	0.0596	106.7008
	IBPSO	**0.0584**	**0.0668**	**0.0025**	**0.0593**	107.3102
Lungcancer	GA	0.0264	0.0283	0.0006	0.0275	**49.9007**
	BFA	0.0271	0.0281	**0.0003**	0.0277	102.6213
	BCS	0.0281	0.0549	0.0113	0.0373	100.4391
	BPSO	0.0271	0.0287	0.0004	0.0280	103.4912
	BBA	0.0269	0.0280	**0.0003**	0.0276	102.6021
	IBPSO	**0.0256**	**0.0269**	0.0004	**0.0264**	107.6533

TABLE 5 Fitness Values Obtained by Different Algorithms (Datasets 9 to 16)

Conclusion: In this paper, the feature selection problem is investigated. First, a joint feature selection problem is formulated, and then we propose an efficient algorithm called IBPSO to solve the formulated problem. In IBPSO, we first introduce the Lévy flight mechanism to improve the local search performance of the algorithm. Second, a weighting inertia coefficient operator is proposed to enhance the global search ability. Moreover, we use the mutation mechanism to improve the population diversity of the algorithm. Finally, a binary method is adopted to make the continuous algorithm suitable for the binary feature selection problem. Experiments are conducted on several classical datasets for the evaluations of the proposed algorithm, and the results show that the overall performance of IBPSO outperforms GA, BFA, BCS, BPSO and BBA for solving the feature selection problem. In our future work, more test datasets will be considered to further evaluate the proposed algorithm.

Submitted By
Name: Rokibul Hasan
ID: MCE 079 05534