A Hybrid Algorithm Based on Binary Chemical Reaction Optimization and Tabu Search for Feature Selection of High-Dimensional Biomedical Data

Chaokun Yan
the School of Computer and Information Engineering, Henan University, Kaifeng 475000, China.

Jingjing Ma
the School of Computer and Information Engineering, Henan University, Kaifeng 475000, China.

Huimin Luo
the School of Information Science and Engineering, Central South University, Changsha 410083, China.

Jianxin Wang
the School of Information Science and Engineering, Central South University, Changsha 410083, China.

Follow this and additional works at: https://tsinghuauniversitypress.researchcommons.org/tsinghua-science-and-technology

Part of the Computer Sciences Commons, and the Electrical and Computer Engineering Commons

Recommended Citation

A Hybrid Algorithm Based on Binary Chemical Reaction Optimization and Tabu Search for Feature Selection of High-Dimensional Biomedical Data

Chaokun Yan, Jingjing Ma, Huimin Luo, and Jianxin Wang*

Abstract: In recent years, there have been rapid developments in various bioinformatics technologies, which have led to the accumulation of a large amount of biomedical data. The biomedical data can be analyzed to enhance assessment of at-risk patients and improve disease diagnosis, treatment, and prevention. However, these datasets usually have many features, which contain many irrelevant or redundant information. Feature selection is a solution that involves finding the optimal subset, which is known to be an NP problem because of the large search space. Considering this, a new feature selection approach based on Binary Chemical Reaction Optimization algorithm (BCRO) and k-Nearest Neighbors (KNN) classifier is presented in this paper. Tabu search is integrated with CRO framework to enhance local search capacity. KNN is adopted to evaluate the quality of selected candidate subset. The results for an experiment conducted on nine standard medical datasets demonstrate that the proposed approach outperforms other state-of-the-art methods.

Key words: feature selection; biomedical data; chemical reaction optimization; tabu search

1 Introduction

With the developments of key technologies in the biomedical and health fields, biomedical data has accumulated (expected to reach 25 exabytes in 2020)[1]. Such large amounts of data cannot be processed directly by experts in a short time for diagnosis or treatment, which gives rise to new requirements for data mining and machine learning[2]. Moreover, many real-world datasets often involve many features, and not all features are essential. The irrelevant features do not only lead to

insufficient classification accuracy, but also cause extra difficulties in finding potentially useful knowledge[3]. Considering the challenges in extracting valuable information and determining the important features of large datasets, feature selection, also known as variable selection or attribute selection, has attracted much interests in the biomedical domain.

The goal of feature selection is to select the most informative features from a given medical dataset. It is used to reduce the dimensionality of data, improve the prediction accuracy, and reduce the computational cost for disease diagnosis. In general, the traditional approaches can be broadly categorized into three: filter, wrapper, and embedded approaches[4]. For the filter model, features are evaluated only based on the general data characteristics, without utilizing any mining algorithm, which is effective in terms of computational cost. However, the main drawback of the filter model is that the dependencies among the features are ignored, and the obtained feature subsets could contain some redundant information, which leads to low classification accuracy. TRank algorithm[5]

● Chaokun Yan and Jingjing Ma are with the School of Computer and Information Engineering, Henan University, Kaifeng 475000, China. E-mail: ckyan@henu.edu.cn; majingjing4729@163.com.
● Huimin Luo and Jianxin Wang are with the School of Information Science and Engineering, Central South University, Changsha 410083, China. E-mail: luohuimin@csu.edu.cn; jxwang@mail.csu.edu.cn.
∗ To whom correspondence should be addressed.
  Manuscript received: 2018-04-06; accepted: 2018-05-04

is a filtering algorithm that has been commonly used in testing the difference of the feature between two groups. Unlike the filtering method, the wrapper model usually selects feature subset while relying on a classifier[6] and can achieve higher classification accuracy, but it involves the repeated training of classifiers, which leads to a high computational cost. Embedded-based feature selection methods are special cases of wrapper methods that are characterized by a deeper interaction between the construction of the learning algorithm and the feature selection. Decision tree is a popular embedded method, in which features are automatically selected based on the class discrimination capability[7]. Comparatively speaking, the wrapper approach utilizes the performance of machine learning algorithms as an evaluation standard to estimate the selected features, which is more flexible.

Feature selection can be formulated as a combinatorial optimization problem and aims to minimize the size of feature subset and achieve optimum classification performance. In the few past years, research on feature selection was mainly focused on two aspects of the procedure: subset search and subset evaluation[8]. The former involves selecting a subset of features based on the corresponding strategy, while the latter involves evaluating the quality of the current selected best feature subset and deciding whether to replace the preselected feature subset. During the search process, an exhaustive search for the best feature subset of a given dataset would be practically impossible and would be impeded by the problem of combinatorial explosion. Compared with exhaustive search based algorithms, branch-and-bound algorithms[9] use monotonic evaluation functions and reduce the time cost; however, it is difficult to design evaluation functions, and high-dimensional biomedical problems cannot be tackled. In recent years, meta-heuristic methods have attracted much attention in feature search because of their ability to find global optimal solutions. For example, simulated annealing approach[10] has been developed and applied to parameter determination and feature selection. A binary genetic algorithm[11] was proposed to reduce the number of features, which can extract one hundred features from a set of images in a public Flavia dataset.

Ghanda and Ahmadi[12] proposed a new model based on Particle Swarm Optimization (PSO) and naive Bayesian classification to diagnose Parkinson's disease. In this model, the optimal training data for naive Bayesian classification were selected using PSO algorithm. Hu et al.[13] presented a nature inspired approach, Improved

Shuffled Frog-Leaping Algorithm (ISFLA), which has been successfully applied to feature selection problems in molecular diagnosis of diseases by introducing a chaos memory weight factor, an absolute balance group strategy, and an adaptive transfer factor. The proposed algorithm has improved the classification accuracy and the efficiency of disease diagnosis. Vieira et al.[14] presented a Modified Binary Particle Swarm Optimization (MBPSO) method for feature selection problems and applied it to mortality prediction in septic patients. The MBPSO was used as a wrapper method to select feature subset for Support Vector Machine (SVM) classification. A Binary Artificial Bee Colony (BABC) algorithm[15] was used to find the optimal feature subset in heart diseases identification, and then k-Nearest Neighbors (KNN) model was utilized to evaluate the selected features. The performance of the BABC algorithm has been validated on Cleveland heart disease dataset. In current studies, wrapper-based methods have been widely used because the wrapper model can find feature subsets more effectively.

Chemical Reaction Optimization (CRO)[16, 17] is a newly proposed chemical reaction-inspired meta-heuristic algorithm with low computational cost and high efficiency, compared to the other meta-heuristic algorithms, and it has been applied to many optimization problems in both discrete and continuous domains. To the best of our knowledge, CRO has not been applied to feature selection studies. In this paper, Binary Chemical Reaction Optimization (BCRO) is employed to address feature selection problems in the biomedical domain. Considering the capacity of Tabu Search (TS) in finding local optimum solutions[18, 19], it is combined with BCRO to realize a better performance. The optimal feature subset is selected using BCRO-TS, and then KNN is used for fitness evaluation based on the selected features. The performance of the proposed method BCRO-TS-KNN is evaluated on nine public biomedical datasets. Furthermore, the impact of two other classifiers[20], namely naive Bayes and SVM, is measured on these benchmark datasets. Experimental results show that our proposed approach results in improvements in the identification of relevant feature subsets and classification performance.

The remainder of this paper is organized as follows. In Section 2, we describe the BCRO algorithm and present the application flowchart of the BCRO-TS framework in feature selection. In Section 3, we present and discuss our experimental results. Finally, in Section 4 we present our conclusion and some potential future work.

2 BCRO-TS-Based Feature Selection

2.1 Basic operations of BCRO algorithm

The BCRO[16] algorithm imitates the chemical reaction process and is governed by the thermodynamics laws. The BCRO process is divided into three stages: initialization, iteration, and finishing. In the initialization stage, all parameters and the population are initialized. In the iteration stage, four elementary reactions are implemented in BCRO, namely decomposition, on-wall ineffective, synthesis, and intermolecular ineffective collision. During the iterations, two parameters, α and β, play important roles in realizing the elementary reactions. The parameter α controls the occurrence of decomposition and on-wall ineffective collision. Decomposition occurs if the number of hits that a molecule has taken is larger than α; otherwise, on-wall ineffective collision occurs. The parameter β controls the occurrence of synthesis and intermolecular ineffective collision. Synthesis occurs if the kinetic energy of two selected molecules are less than β; otherwise, intermolecular ineffective collision is triggered. The optimum solution will be output in the final stage. A molecule has two kinds of energies, Potential Energy (PE) and Kinetic Energy (KE). The potential energy quantifies the molecular structure in terms of energy and we utilize it to construct the objective function for evaluating the corresponding solution. The kinetic energy allows one molecule to move to a higher potential state. Therefore, the kinetic energy of a molecule represents its ability of escaping from a local minimum. In this study, each molecule in the chemical reaction is modeled as a binary string to represent a solution. The search process is terminated when the optimal solution converges. The four basic reactions BCRO are implemented as follows.

On-wall ineffective collision is used to discover a neighbor of solution ω in a search space, which is known as the mutation process. In the solution ω, 0 indicates that the corresponding item is in the molecular structure; otherwise, it indicates the opposite condition. The mutation operator replaces a random position $\omega(i)$ with a binary number 0 or 1, which can be represented as follows:

$$\omega : [1,0,\mathbf{0},1,0,0] \rightarrow \omega' : [1,0,\mathbf{1},1,0,0].$$

If $\mathrm{PE}_{\omega'} \leqslant \mathrm{PE}_\omega + \mathrm{KE}_\omega$, then ω' has a better molecular structure and it would replace ω.

Decomposition produces two solutions from one original solution, which is designed according to the half-total-exchange operator that is used to solve the channel assignment problem. This operator duplicates the original molecule to obtain two new molecules and randomly changes the half of each new molecule with 0 and 1, which can be described as follows:

$$\omega : [1,0,0,0,1,0] \rightarrow \begin{cases} \omega_1' : [1,\mathbf{1},0,\mathbf{1},\mathbf{0},0], \\ \omega_2' : [\mathbf{0},0,\mathbf{1},0,1,\mathbf{1}]. \end{cases}$$

If $\mathrm{PE}_\omega + \mathrm{KE}_\omega \geqslant \mathrm{PE}'_{\omega_1} + \mathrm{PE}'_{\omega_2}$, then ω_1' and ω_2' would be conserved in the population, and ω is destroyed; otherwise, ω_1' and ω_2' are destroyed.

Intermolecular ineffective collision is an elementary reaction involving more than one molecule. The effect of energy change of the molecules is similar to that in the on-wall ineffective collision. Suppose the original molecular structures are ω_1 and ω_2, the difference between the molecules' energy determines whether the new generated molecules would replace the original ones. If the new molecules have a lower energy, two new molecular structures ω_1' and ω_2' are obtained from the ω_1 and ω_2 neighborhoods, respectively. This reaction can be formulated as follows:

$$\omega_1 : [1,0,0,1,0,1], \quad \omega_2 : [0,1,0,0,1,0],$$
$$\downarrow$$
$$\omega_1' : [1,\mathbf{1},0,1,0,1], \quad \omega_2' : [0,1,0,\mathbf{1},1,0].$$

Different from the above reactions, the synthesis reaction is a global search to generate a new molecular structure ω' by combining two existing solutions ω_1 and ω_2. Half of the new generated molecule ω' is duplicated from ω_1 with the items in the corresponding position, and the other are derived from ω_2, which can be described as follows:

$$\begin{cases} \omega_1 : [\mathbf{0},\mathbf{1},1,0,\mathbf{0},1], \\ \omega_2 : [1,0,\mathbf{1},\mathbf{0},1,\mathbf{0}] \end{cases} \rightarrow \omega' : [0,1,1,0,0,0].$$

If $\mathrm{PE}_{\omega_1} + \mathrm{PE}_{\omega_2} + \mathrm{KE}_{\omega_1} + \mathrm{KE}_{\omega_2} \geqslant \mathrm{PE}_{\omega'}$, ω' is added to the population, and ω_1 and ω_2 are destroyed. If not, the molecule ω' is destroyed.

2.2 BCRO-TS for feature selection

In BCRO, each molecule can be regarded as a candidate solution of the optimization problem. Each molecule has potential and kinetic energies, which are updated by one of the above mentioned chemical reaction. To avoid cycling through solutions in a search, TS adopts a tabu list to store the forbidden items, which have been recently

visited or limited by users-provided rules. In addition, the algorithm searches the neighbors of solutions to obtain a new potential solution. The solutions in the tabu list will not be chosen. The purpose of the tabu list is to prevent looping in the recent solutions and to diversify the search of the search space. In this paper, tabu list and neighborhood searches are combined with BCRO to realize a better search performance. For algorithm BCRO-TS, once one of the four elementary reactions has been performed, the best solution in the iteration would be checked, and tabu search is employed to search neighbors, which is a local search process. This section introduces the implementation process of the proposed BCRO-TS for feature selection.

Step 1 Initialization. Randomly initialize a population containing M molecules.
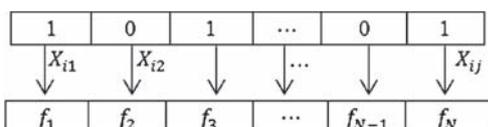
It is common to use binary encoding in feature selection. To initialize a population, we randomly produce M molecules. Each molecule X_i is represented as a binary one-dimensional array of length N. For binary variable X_{ij}, i represents the molecule index, j represents the dimension of this molecule, and $X_{ij} \in \{0,1\}$. The variable X_{ij} corresponds to the input feature f_j, where $j = 1, 2, \ldots, N$. If feature f_j is selected then $X_{ij} = 1$; otherwise, $X_{ij} = 0$. The process is depicted in Fig. 1.

Step 2 Search mechanism. The best feature subset with the smallest potential energy is selected by the BCRO algorithm.

The aim of the selection mechanism in BCRO is to constantly improve the molecules (solution candidates) over all fitness values, which are the opposite of the PE values. The selection mechanism helps BCRO discard bad molecules and keep the best individuals. For each molecule X_i, the feature f_j is selected randomly in the following Formula (1).

$$f_j = \begin{cases} 1, & r < pro; \\ 0, & r \geqslant pro \end{cases} \qquad (1)$$

Here r is a random number in $(0, 1)$, and pro is a threshold parameter representing the probability of selecting the relevant features. After a series of selection operations, BCRO converges to an optimal subset of features in an iteration of a binary strategy. Figure 2 gives a detailed description of one on-wall ineffective collision reaction.

Step 3 Fitness evaluation. Use the evaluation function to calculate the value for each individual and identify the global optimum molecule.

Feature selection involves two main objectives: to maximize the classification accuracy and minimize the number of features. In this study, we use the proposed meta-heuristic method BCRO-TS to select the optimal feature subset, and use the KNN-based fitness function defined in Ref. [13] for evaluation.

$$acc(KNN) = \frac{num_c}{num_c + num_i} \times 100\% \qquad (2)$$

$$fitness = \omega_1 \times acc(KNN) + \omega_2 \times \left(1 - \frac{n}{N}\right) \qquad (3)$$

Here, ω_1 and ω_2 are set to 1 and 0.001, respectively, as in Ref. [13]. The function $acc(KNN)$ is the classification accuracy based on KNN. N is the total number of features, and n is the number of selected features. The numbers of correctly and incorrectly classified instances are indicated by num_c and num_i, respectively.

The objective function is to identify the significant features and discard irrelevant or redundant features from the original set of features. The function acc defined in Eq. (2) denotes the percentage of correctly classified instances. As defined by Eq. (3), the fitness function has two predefined weight parameters: ω_1 (the classification accuracy) and ω_2 (the selected features). This fitness function can be utilized to obtain a good trade-off between the classification performance and the number of the selected features.

Step 4 Update solution. The best solution is updated

Fig. 1   Schematic of X_{ij} representing the selection of the corresponding feature f_j.
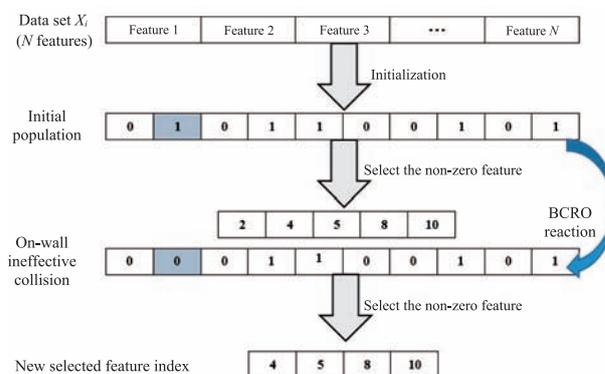
Fig. 2   On-wall ineffective collision of a molecule during iteration.

by TS, which has a strong local search ability to search the neighbors of the optimal solution.

Tabu search is a local neighborhood search algorithm that simulates the optimal characteristics of human memory functions. Tabu search involves a local search combined with a tabu mechanism. At each iteration, the algorithm searches the neighborhood of the best solution to obtain a new one with an improved functional value. After the chemical reaction is completed, the optimum solution $bestsol$ in the population is obtained, and then a TS is employed to search its neighborhoods and update $bestsol$ by a molecule with better value fitness.

Step 5 Continue the iterative process. Return to Step 2 if stopping criteria have not been met.

The description of BCRO-TS algorithm for feature selection is shown as Algorithm 1.

In the BCRO-TS algorithm, parameter $Popsize$ represents the population size, $KELossRate$ denotes the kinetic loss rate during the reaction, $MoleColl$ decides whether intermolecular collision would occur, $buffer$ stores the energy transformed from a portion of kinetic energy, and $InitialKE$ is the molecule's initial kinetic energy. Other necessary parameters are as follows: $neighborMol(\omega)$ randomly changes an item of ω and returns a new molecular structure; $tabuTableUpdate(\omega)$ adds molecular structure ω to the tabu table, which is an FIFO data structure; $judge(\omega)$ is used to determine whether solution ω is in the tabu list; if ω is not in the tabu list, then return 0; otherwise, return 1. Parameters $tabuLength$ is the length of the tabu table, and $numNeighbor$ is the number of the $bestSol$'s neighbors. $fitness(\omega)$ is used to calculate the fitness value of molecule ω, which is the opposite of the PE value.

3   Results and Discussion

To verify the effectiveness of the proposed method, our experimental results consist of two parts. First, the performance of the proposed method is compared with those of other state-of-the-art approaches on nine biomedical datasets. Then the impact of three classifiers on CRO-TS is evaluated to test the robustness of the proposed approach.

3.1   Datasets

The experiment was conducted on several well-known and recognized biomedical datasets which were downloaded from the Kent Ridge biomedical dataset repository at http://leo.ugr.es/elvira/DBCRepository/. These datasets include ALL-AML-Leukemia, ColonTumor, and Nervous-

---

Algorithm 1 BCRO-TS Algorithm for Feature Selection

Input: Problem-specific information (the objective function f, instances M, and the dimensions of the dataset N)

Output: the optimal solution

1: Assign parameter values to $Popsize$, $KELossRate$, $MoleColl$, and $InitialKE$
2: Randomly initialize populations
3: **while** (the stopping criteria not met) **do**
4:     Get t randomly in interval $[0, 1]$
5:     **if** $t > MoleColl$ **then**
6:         Select a molecule ω from Pop randomly
7:         **if** decomposition criterion met **then**
8:             $(\omega'_1, \omega'_2, Success)$= decompose$(\omega, buffer)$
9:             **if** $Success$ **then** remove ω from Pop
10:                 Add ω'_1 and ω'_2 to Pop
11:             **end if**
12:         **else**
13:             ineff_coll_on_wall $(\omega, buffer)$
14:         **end if**
15:     **else** // $t \leqslant MoleColl$
16:         Select molecules ω_1, ω_2 from Pop
17:         **if** synthesis criterion met **then**
18:             $(\omega', Success)$=synthesis(ω_1, ω_2)
19:             **if** $Success$ **then** remove ω_1, ω_2 from Pop
20:                 Add ω' to Pop
21:             **end if**
22:         **else**
23:             inter_ineff_coll(ω_1, ω_2)
24:         **end if**
25:     **end if** // $t > MoleColl$
26:     evaluate each molecule by KNN and compute fitness
27:     obtain the current optimal solution $bestSol$ with maximum fitness value
28:     //update $bestSol$ by TS
29:     **while** $itr < numNeighbor$ **do**
30:         construct the neighbor ω' of $bestSol$
31:         **if** judge(ω') **then**
32:             **if** $fitness(\omega') > fitness(bestsol)$ **then**
33:                 update $bestSol$ with ω'
34:                 $tabuTableUpdate(\omega')$
35:             **end if**
36:         **end if**
37:     **end while**
38: **end while**
39: Get the current optimal solution $bestSol$

System, which provide data relating to gene expression, protein profiling, and genomic sequence for classification and disease diagnosis. All data in these datasets are high-dimensional and may include irrelevant or weak correlation features. Moreover, the dimensional scopes of these datasets are from 2000 to 12 600. The nine typical high-dimensional biomedical datasets used in our study are listed in Table 1.

Table 1   Benchmark datasets.

Dataset	Number of instances	Number of attributes	Number of classes
ALL-AML_train	38	7130	2
ColonTumor	62	2000	2
DLBCLOutcome	58	7129	2
DLBCL-Stanford	47	4026	2
lungCancer_train	32	12 534	2
LungCancer-Ontario	39	2880	2
NervousSystem	60	7129	2
DLBCL-NIH-train	160	7400	2
LungCancer-Harvard1	203	12 600	5

3.2   Parameter setting

To decide the optimal combination of BCRO-TS parameters, an orthogonal experimental design[21, 22] was conducted. We designed an orthogonal table L_{16} (4^5) to select the optimal parameters used in our study. The parameters in BCRO-TS are reported in Table 2. To assess the performance of the BCRO-TS method, we compare it with the basic BCRO and other three state-of-the-art methods: ISFLA[13] combined with KNN, MBPSO combined with SVM (MPSO-SVM)[14], and GA[11], which are designed for feature selection of biomedical data.

A 10-fold cross validation was conducted ten times to test the performance of all approaches. In each cross validation, the instances were randomly divided into ten parts. Each part was taken as a test set in turn with the remaining nine parts as train set. The accuracy and the number of selected features were used to evaluate the performance of these methods. To maintain the fairness of the experiment, the experiments of all methods were repeated ten times, and we took the mean values as the final results.

3.3   Experimental results and analysis

In this section, we present the experimental results of

Table 2   Parameter settings.

Parameter	BCRO	BCRO-TS
Popsize	20	20
KELossRate	0.2	0.2
InitialKE	5000	5000
MoleColl	0.2	0.2
α	1500	1500
β	10	10
Buffer	0	0
tabuLength	–	25
numNeighbor	–	100

BCRO-TS and other approaches on benchmark datasets and conduct performance analysis. Furthermore, the effects of three classifiers on the proposed method are evaluated.

3.3.1   Algorithm comparison

For each algorithm, there are five attributes tabulated: (1) the average accuracy (Acc), (2) the standard deviation (std), (3) the average number of feature subsets ($AvgN$), (4) the highest accuracy (Max), and (5) the lowest accuracy (Min).

From Table 3, it can be seen that compared with the other algorithms, the results produced by the BCRO-TS algorithm achieved the best average accuracy (Acc) on the nine datasets. For datasets ALL-AML_train, lungCancer_train, and LungCancer-Harvard1, the average accuracy obtained by the BCRO-TS were 99.918%, 99.875%, and 93.475%, respectively. For the other six benchmark datasets, the average accuracy achieved by our proposed algorithm is far greater than those obtained by ISFLA, GA, and MPSO. In addition to the high performance, the robustness is an important factor in evaluating a classifier. The standard deviations of all criteria for BCRO-TS in almost all datasets were small. The smaller the standard deviation, the more stable the experimental results. The standard deviation of the BCRO-TS average accuracy was the smallest for all the datasets excluding LungCancer-Ontario, NervousSystem, and LungCancer-Harvard1 datasets.

Compared with ISFLA, GA, and MPSO feature selection methods, the BCRO-TS had the smallest number of features ($AvgN$) for all datasets excluding LungCancer-Ontario and LungCancer-Harvard1. There is no conclusive evidence that the accuracy will increase with the use of more features. As shown in Table 3, the ISFLA had the maximum $AvgN$ of 50.7 and 27.8 for lungCancer_train and LungCancer-Harvard1 datasets, respectively, but corresponding accuracies were not the largest.

Table 3 shows a list of the average accuracy when selecting the best combination of features. Our proposed BCRO-TS algorithm obviously outperformed the other competitive methods in terms of accuracy, which shows that BCRO-TS has a better feature selection capacity for biomedical datasets, especially for ColonTumor, DLBCLOutcome, DLBCL-Stanford, LungCancer-Ontario, NervousSystem, and DLBCL-NIH-train. For datasets ALL-AML_train and lungCancer_train, the standard deviation of BCRO-TS is less than 1 which further proves the good robustness of the algorithm.

**Table 3   The experiment results of four algorithms on the benchmark datasets.**

Dataset	Algorithm	Acc (%)	std of Acc	AvgN	std of AvgN
ALL-AML_train	**BCRO-TS**	99.918	0.724	26.4	2.9
	GA	94.263	2.23	34.1	5.82
	ISFLA	96.342	2.632	35.9	4.86
	MPSO	91.552	3.137	41.4	6.13
ColonTumor	**BCRO-TS**	93.123	1.242	33.33	4.8
	GA	85.223	1.988	39.8	5.06
	ISFLA	89.565	1.984	37.1	5.87
	MPSO	80.064	2.872	42.1	7.09
DLBCLOutcome	**BCRO-TS**	81.09	1.501	25.1	4.72
	GA	67.017	2.842	31	4.24
	ISFLA	72.103	1.966	29.1	5.97
	MPSO	63.483	3.894	30.7	4.1
DLBCL-Stanford	**BCRO-TS**	97.191	1.973	12.5	3.69
	GA	86.234	3.875	15.2	3.89
	ISFLA	90.213	2.596	15.6	5.6
	MPSO	84.957	3.252	18.1	3.94
lungCancer_train	**BCRO-TS**	99.875	0.25	37.4	5.57
	GA	96.125	1.865	50	8.64
	ISFLA	97.656	2.094	50.7	10.1
	MPSO	96.25	2.86	47.9	5.869
LungCancer-Ontario	**BCRO-TS**	93.487	3.889	11.5	3.32
	GA	77.410	3.579	10.7	4.20
	ISFLA	83.744	3.889	13.6	2.4
	MPSO	79.821	6.392	15	3.38
NervousSystem	**BCRO-TS**	82.717	2.682	24.8	3.6
	GA	70.7	3.225	36.2	4.07
	ISFLA	75.3667	2.333	30.4	7.4
	MPSO	69.05	2.575	29.7	3.95
DLBCL-NIH-train	**BCRO-TS**	72.006	1.285	27.9	4.3
	GA	64.681	1.821	31.6	4.2
	ISFLA	67.138	3.519	34	6
	MPSO	60.863	2.650	32.2	4.81
LungCancer-Harvard1	**BCRO-TS**	93.457	1.351	26.6	3.88
	GA	84.626	0.766	20.3	6.07
	ISFLA	88.020	0.685	27.8	9.1
	MPSO	86.271	1.673	25.2	8.82

As a wrapper strategy, BCRO-TS is combined with KNN classification to implement feature selection. Meanwhile, the binary approach formulates the feature selection problem as a function optimization problem to obtain an optimal feature set with optimal average accuracy and optimal number of features. To further test the impact of the tabu search, we compare BCRO-TS with BCRO (i.e., without TS). As shown in Figs. 3 and 4, BCRO-TS achieved better performance using fewer features, which demonstrates that tabu search can improve the local search capability of BCRO.

The aim of feature selection is to reduce the dimensionality of the original data and improve the efficiency of the search mechanism. In addition, the feature selection process also requires considerable execution time. The running time of the proposed algorithm depends on both the convergence ability of the algorithm and the scale of datasets. Figure 5 shows the time cost of BCRO-TS and the other algorithms. As we can see from Fig. 5, CRO algorithm achieved a better performance on all benchmarks except for ColonTumor, DLBCL-NIH-train, and LungCancer-Harvard1. BCRO-TS was slightly worse
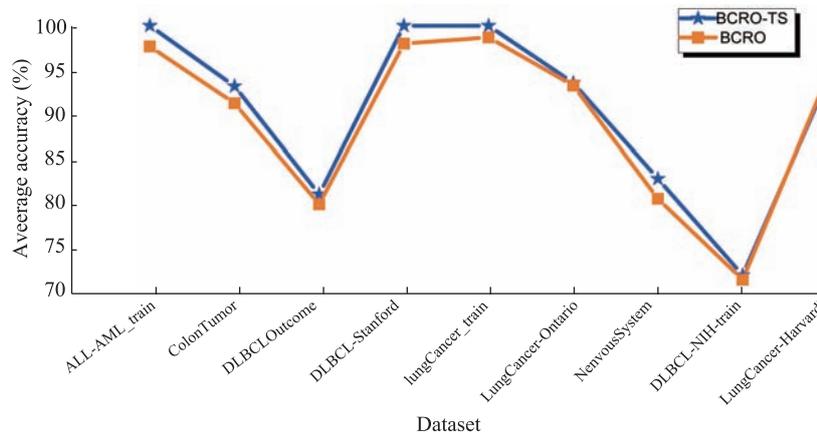
Fig. 3   Average accuracy obtained by BCRO-TS and BCRO.
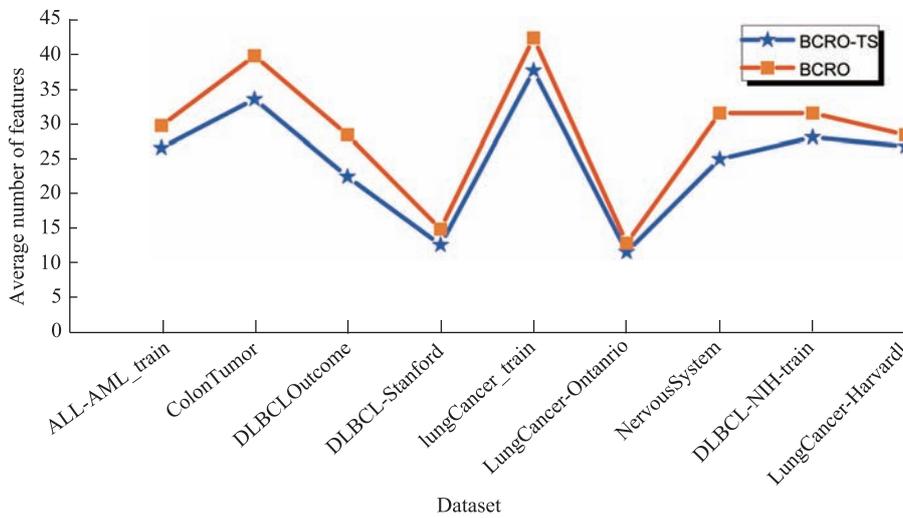
Fig. 4   Average number of features selected by BCRO-TS and BCRO.
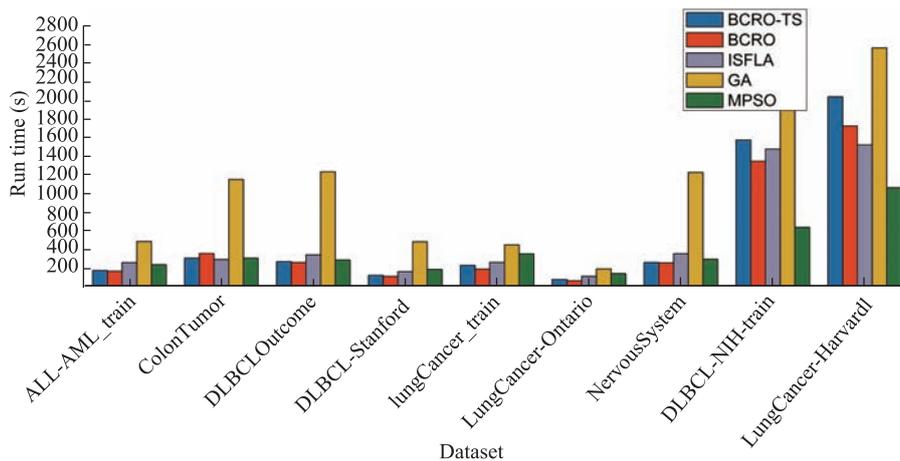
Fig. 5   Running time of all compared algorithms.

than BCRO algorithm in terms of execution time because of the additional search operation caused by the tabu search. However, BCRO-TS still performed better than ISFLA, GA, and MPSO on six datasets.

3.3.2   Impact of the three classifiers

As mentioned above, BCRO-TS achieved good classification performance for disease diagnosis. KNN classifier was used to evaluate the feature subsets selected

by BCRO-TS. The impact of the classifier on the BCRO-TS performance is not clear. Based on nine benchmark datasets, the impact of three popular classifiers, KNN, SVM, and Naive Bayes (NB) were evaluated, and the results in terms of accuracy and the number of selected features are reported in Table 4. The 10-fold cross validation experiments were performed to evaluate the classifier model. It can be seen that the performances of KNN and SVM classifiers for six datasets ALL-AML_train, ColonTumor, lungCancer_train, LungCancer-Ontario, NervousSystem, and DLBCL-NIH-train are very close. According to the results, we can conclude that KNN-based BCRO-TS has better robustness for feature selection. As can be seen from Table 4, the average accuracy and the average number of feature subsets obtained by BCRO-TS with KNN outperformed those obtained by BCRO-TS with the other two classifiers.

4 Conclusion

Feature subset selection is a fundamental technique in many applications, and different evolutionary algorithms have been developed to solve different feature selection problems. However, the increase of dimensionality of the used data poses a severe challenge to many existing feature selection methods with respect to efficiency and effectiveness. In this study, BCRO was first used to solve the feature selection problem. Then, tabu search algorithm was combined with BCRO to design a hybrid algorithm BCRO-TS, which can efficiently solve feature selection problem for high-dimensional biomedical data. Moreover, KNN classifier was used for fitness evaluation based on the selected features, and it exhibited better performance than SVM and NB classifiers. Experiment results show that BCRO-TS-KNN can use fewer features and achieve a higher classification accuracy simultaneously. For most

Table 4 Impact of three classifiers on BCRO-TS.

Dataset	Classifier	Max (%)	Min (%)	Avg (%)	$AvgN$
	KNN	100	97.632	99.918	26.4
ALL-AML_train	SVM	100	96.423	99.231	27
	NB	97.631	87.894	91.339	29.11
	KNN	95.304	88.356	93.123	33.33
ColonTumor	SVM	93.828	87.194	92.184	35.88
	NB	93.163	85.905	88.037	35.11
	KNN	89.962	76.582	81.09	25.1
DLBCLOutcome	SVM	88.031	79.927	82.925	27.11
	NB	74.927	70.444	73.011	29.77
	KNN	100	96.376	97.191	12.5
DLBCL-Stanford	SVM	97.617	89.106	95.147	13.44
	NB	94.106	85.191	89.314	15
	KNN	100	95.937	99.875	37.44
lungCancer_train	SVM	100	97.876	99.375	38.88
	NB	99.475	91.875	96.905	42.88
	KNN	99.179	90.513	93.487	11.5
LungCancer-Ontario	SVM	94.615	87.179	90.797	8.77
	NB	92.664	85.228	89.530	11.77
	KNN	89.513	80.667	82.717	24.8
NervousSystem	SVM	87.166	80.166	84.283	29.88
	NB	85.433	76.266	81.489	26.89
	KNN	76.912	69.600	72.006	27.9
DLBCL-NIH-train	SVM	75.125	60.187	70.02	27.11
	NB	74.625	69.524	71.370	30.44
	KNN	96.404	90.857	93.457	26.6
LungCancer-Harvard1	SVM	98.129	92.272	96.084	28.56
	NB	89.251	76.110	85.925	26.11

biological datasets, BCRO-TS-KNN can achieve high performance with a very small number of features in a short time when compared with other art-of-the-state methods.   The proposed algorithm can be a useful preprocessing tool in selecting informative features from high-dimensional biomedical data, as well as in mining functions of biological data in disease diagnosis and improving the efficiency of disease diagnosis.   In the future, we will further improve the exploration and exploitation of BCRO by integrating it with other local search strategies or swarm intelligent algorithms.   In addition, more classifiers will be evaluated to improve the performance of the wrapper strategy.

Acknowledgment

References

[1]   K. Lee, Z. H. Man, D. H. Wang, and Z. W. Cao, Classification of microarray datasets using finite impulse response extreme learning machine for cancer diagnosis, presented at the IECON 2011—37th Annu. Conf. of the IEEE Industrial Electronics Society, Melbourne, Australia, 2011.

[2]   H. Liu and Z. Zhao, Manipulating data and dimension reduction methods: Feature selection, presented at the Encyclopedia of Complexity and Systems Science, Berlin, Germany, 2015.

[3]   Z. M. Hira and D. F. Gillies, A review of feature selection and feature extraction methods applied on microarray data, *Advances in Bioinformatics*, vol. 2015, p. 198363, 2015.

[4]   K. Kira and L. A. Rendell, The feature selection problem: Traditional methods and a new algorithm, in *Proc. Tenth National Conf. Artificial Intelligence*, San Jose, CA, USA, 1992, pp. 129–134.

[5]   P. Baldi and A. D. Long, A Bayesian framework for the analysis of microarray expression data: Regularized t-test and statistical inferences of gene changes, *Bioinformatics*, vol. 17, no. 6, pp. 509–519, 2001.

[6]   N. Verbiest, J. Derrac, C. Cornelis, S. García, and F. Herrera, Evolutionary wrapper approaches for training set selection as preprocessing mechanism for support vector machines: Experimental evaluation and support vector analysis, *Applied Soft Computing*, vol. 38, pp. 10–22, 2016.

[7]   Y. H. Peng, Z. Q. Wu, and J. M. Jiang, A novel feature selection approach for biomedical data classification,

[8]   *Journal of Biomedical Informatics*, vol. 43, no. 1, pp. 15–23, 2010.

[8]   B. Xue, M. J. Zhang, W. N. Browne, and X. Yao, A survey on evolutionary computation approaches to feature selection, *IEEE Transactions on Evolutionary Computation*, vol. 20, no. 4, pp. 606–626, 2016.

[9]   M. Yagiura and T. Ibaraki, On metaheuristic algorithms for combinatorial optimization problems, *Systems and Computers in Japan*, vol. 32, no. 3, pp. 33–55, 2001.

[10] S. W. Lin, Z. J. Lee, S. C. Chen, and T. Y. Tseng, Parameter determination of support vector machine and feature selection using simulated annealing approach, *Applied Soft Computing*, vol. 8, no. 4, pp. 1505–1512, 2008.

[11] O. H. Babatunde, L. Armstrong, J. S. Leng, and D. Diepeveen, A genetic algorithm-based feature selection, *International Journal of Electronics Communication and Computer Engineering*, vol. 5, no. 4, pp. 899–905, 2014.

[12] N. K. Ghanad and S. Ahmadi, Combination of PSO algorithm and naive Bayesian classification for Parkinson disease diagnosis, *Advances in Computer Science*: *An International Journal*, vol. 4, no. 4, pp. 119–125, 2015.

[13] B. Hu, Y. Q. Dai, Y. Su, P. Moore, X. W. Zhang, C. S. Mao, J. Chen, and L. X. Xu, Feature selection for optimized high-dimensional biomedical data using the improved shuffled frog leaping algorithm, presented at the IEEE/ACM Trans. on Computational Biology and Bioinformatics, 2016.

[14] S. M. Vieira, L. F. Mendonça, G. J. Farinha, and J. M. C. Sousa, Modified binary pso for feature selection using SVM applied to mortality prediction of septic patients, *Applied Soft Computing*, vol. 13, no. 8, pp. 3494–3504, 2013.

[15] B. Subanya and R. R. Rajalaxmi, Feature selection using artificial bee colony for cardiovascular disease classification, presented at the 2014 Int. Conf. on Electronics and Communication Systems (ICECS), Coimbatore, India, 2014.

[16] B. Xing and W. J. Gao, Chemical-reaction optimization algorithm, in *Innovative Computational Intelligence: A Rough Guide to 134 Clever Algorithms*. Springer, 2014.

[17] B. Alatas, ACROA: Artificial chemical reaction optimization algorithm for global optimization, *Expert Systems with Applications*, vol. 38, no. 10, pp. 13170–13180, 2011.

[18] F. Glover and M. Laguna, Tabu search, in *Handbook of Combinatorial Optimization*, D. Z. Du and P. M. Pardalos, eds. Springer, 2013, pp. 3261–3362.

[19] F. Glover, Tabu search: A tutorial, *Interfaces*, vol. 20, no. 4, pp. 74–94, 1990.

[20] B. S. C. Wade, S. H. Joshi, B. A. Gutman, and P. M. Thompson, Machine learning on high dimensional shape data from subcortical brain surfaces: A comparison of feature selection and classification methods, *Pattern*

Recognition, vol. 63, pp. 731–739, 2017.

[21] L. Franek and X. Y. Jiang, Orthogonal design of experiments for parameter learning in image segmentation, *Signal Processing*, vol. 93, no. 6, pp. 1694–1704, 2013.

[22] W. G. Cui, X. H. Li, S. B. Zhou, and J. Weng, Investigation on process parameters of electrospinning system through orthogonal experimental design, *Journal of Applied Polymer Science*, vol. 103, no. 5, pp. 3105–3112, 2007.

Chaokun Yan received the PhD degree from Central South University in 2013. He is an associate professor in School of Computer and Information Engineering, Henan University, Kaifeng, China. His research interests include data mining and bioinformatics.

Huimin Luo is working toward the PhD degree in the School of Information Science and Engineering, Central South University. Her current research interests include bioinformatics and systems biology.

Jianxin Wang received the BEng and MEng degrees from Central South University, China, in 1992 and 1996, respectively, and the PhD degree in computer science from Central South University, China, in 2001. He is the vice dean and a professor in School of Information Science and Engineering, Central South University, Changsha, China. His current research interests include algorithm analysis and optimization, parameterized algorithm, bioinformatics, and computer network. He is a senior member of the IEEE.

Jingjing Ma received the bachelor degree from Henan Normal University in 2016. She is now a master student in the School of Computer and Information Engineering, Henan University. Her research interests are bioinformatics, data mining, and deep learning.