Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 154 (2019) 274–282

Procedia Computer Science

www.elsevier.com/locate/procedia

8th International Congress of Information and Communication Technology, ICICT 2019

A Hybrid Cancer Classification Model Based Recursive Binary Gravitational Search Algorithm in Microarray Data

Xiao Hong Han[a,] *, Deng Ao Li[a], Li Wang[a]

[a]*Taiyuan Universicy of Technology, No 79 yingze west street, taiyuan,030024,China*

Abstract

Nowadays, in clinical medicine diagnosticians usually use DNA microarray datasets for diagnosis and classification of cancer. However, DNA microarray datasets typically have very large number of genes and less number of samples, therefore, before diagnosis and classification of cancer it is quite requisite to select most relevant genes. In this paper, we have developed a two phase classification model in which most relevant genes are selected by integrating ReliefF with Recursive Binary Gravitational Search Algorithm (RBGSA) in the help of a classifier of Multinomial Naive Bayes. The RBGSA recursively transforms a very raw gene space to an optimized one at each iteration while not degrading the accuracy. We evaluate our model by comparing it with 6 other known methods on 6 different microarray datasets of cancer. Comparison results show that our model gets substantial improvements in accuracy over other methods.

© 2019 The Authors. Published by Elsevier Ltd.
This is an open access article under the CC BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/4.0/)
Selection and peer-review under responsibility of the 8th International Congress of Information and Communication Technology, ICICT 2019.

Keywords: classification; Gravitational Search Algorithm; Microarray data;

1. Introduction

Microarray cancer data [1] has been widely and successfully applied to cancer classification research in biomedicine. These types of cancer datasets usually have the characteristics of larger number of gene and less number of samples, in other words, the number of gene is much larger than the number of samples. Among these

* Corresponding author. Tel.: +134-8532-5973; fax: +134-8532-5973.
E-mail address: hanxiaohong@tyut.edu.cn

1877-0509 © 2019 The Authors. Published by Elsevier Ltd.
This is an open access article under the CC BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/4.0/)
Selection and peer-review under responsibility of the 8th International Congress of Information and Communication Technology, ICICT 2019.
10.1016/j.procs.2019.06.041

hundreds of thousands of gene expressions only a few genes are significant and useful for cancer classification [2]. Therefore, feature (gene) selection is very important for the establishment of an effective classification model. Feature selection in general can be grouped into two types according to whether a certain learning algorithm is integrated into the process of feature selection or not. If feature selection does not depend the learning algorithm, this type of feature selection is called a filter approach. Otherwise, it is a wrapper approach.

Filter methods include univariate filters and multivariate filters. Univariate filters search and evaluate each gene separately by surveying its inherent natures with regard to discriminate class, thus leading to unreliable outcomes because of not considering gene interactions. While multivariate filters search and evaluate the subset of genes through surveying their inherent natures with regard to different classes, which can promise better results than univariate filters in identifying the most relevant genes in microarray data. Relief is one of the multivariate filter approaches [3,4] based property ranking scheme. Kononenko later developed an improved method called ReliefF based on Relief [5]. In many classification tasks, Relief and ReliefF are usually used as pre-processing approaches for feature selection prior to the model learning. These types of methods not only are effective but also are able to accurately assess the importance of properties [6].

Wrapper approaches evaluate the quality of a feature subset through classification accuracy. These types of algorithms heuristically look for the important gene set within an exponential search space. Backward Feature Elimination (BFE), Ant Colony Optimization, Particle Swarm Optimization (PSO), and Genetic Algorithm (GA) [7-9] belong to wrapper approaches

GSA is a global search approach in which the Newton's Second Law of gravity is used to search for an optimal solution for optimization problems. GSA looks for optimal feature subsets through a fitness function value and is able to obtain fast convergence toward a global optimum within limited iterative times [10,11]. Currently, there are lots of wrapper methods based on GSA proposed to select important features in real applications [12–15].

For microarray datasets, the filter approach usually has the advantage of costing less computational burden. Wrapper approaches have the issue of highly computational overhead in assessing candidate feature subsets, since wrapper methods use a certain learning algorithm on the dataset for each feature subset [16]. Because of the embedding of a certain learning algorithm, wrapper approaches are able to get better performance as regards to classification accuracy than filter methods [17].

Recursive feature selection approaches have gained much more attention[18-20]. These methods recursively eliminated gene features from microarray cancer data. It is observed that these methods improve the overall accuracy slightly and reduce the computational cost using less features, however, they lose the optimal solution.

In this work, we develop a novel classification model called ReliefF -RBGSA-MNB(ReliefF- Recursive Binary Gravitational Search Algorithm-Multinomial Naive Bayes) to accurately classify test data by selecting most informative and discriminative gene subset. This model integrates ReliefF and RBGSA into an united approach thus simultaneously obtaining less computational cost and higher classification performance. In our wrapper method, we develop a recursive binary GSA(RBGSA) scheme motivated by [18-21] which gradually transforms a very row gene set to an optimized one through decreasing the gene set at each iteration. The RBGSA selects important features while not spoiling the accuracy and simultaneously reduces the computational cost.

In Section 2, we present the proposed ReliefF-RBGSA-MNB model in detail. Then the experimental setup and results are given in Section 3. Section 4 provide s conclusions.

2. Proposed cancer classification

2.1.The proposed ReliefF-RBGSA-MNB

We use a multivariate filter to remove redundant and irrelevant genes. Before gene pre-filtering, we first perform a data preprocessing step, we use average values to replace the losing values of gene data and all data is normalized using Eq.(1) .

$$e_{new} = \frac{e - \mu}{\sigma} \tag{1}$$

in which $\mu = 0$ and $\sigma = 1$. e_{new} is the converted value of e.

Then we use the ReliefF to reduce the high dimensionality of gene space. We denote the training set with m samples as $\boldsymbol{D}$= { (x_i^n, y_i)| $x_i^n \in \boldsymbol{X}$, $y_i \in L$, $1 \le i \le m$ }, and use $\boldsymbol{X}$ to denote the original gene set which has m samples of gene expressions, and each sample x_i^n (i= 1, 2, . . ., m) consists of n features of gene. We use ReliefF to explore **X** to get a small subset $\vec{\mathbf{X}}$, whose dimensionality is s ($s \le n$), in $\vec{\mathbf{X}}$, each sample x_i^s consists of s features of gene. Given a randomly selected instance R from the training set D, k closest neighbour samples (called H_j) are obtained from the identical class and also another k closest neighbour samples are got from other classes (called $M_j(C)$). The weight of feature is used to distinguish different classes. The weight of each feature is calculated through Eq.(2):

$$W[A] = W[A] - \sum_{j=1}^{k} diff(A,R,H_j)/(m \cdot k) + \sum_{C \neq class(R_i)} \left[\frac{P(C)}{1 - P(class(R_i))} \sum_{j=1}^{k} diff(A,R_i,l \right. \quad (2)$$

in which function $diff(A,R_1,H_1)$ represents the discrepancy between the instance R_1 and the instance H_1. We present the ReliefF algorithm in Algorithm 1.

Algorithm 1: The ReliefF algorithm

Input: the training set D, the number of selecting sample m, the number of the nearest neighbors k, the threshold of weights δ.
Output: the set of weights T.

```
1. Set all weights W[A]:=0; T=∅;
2. for i:=1 to m do
3.   Select an instance R from D randomly;
4.   Find k nearest neighbors (Hj) and k nearest neighbors from other classes (Mj(C));
5.   for A:=1 to all attributes do
6.       Update W[A] according to rule (4);
7.       if W[A]≥ δthen
8.           Add W[A]to T;
9.   end for
10. end for
11. Generate X⃗according to T.
```

2.2. Gene Optimization with recursive BGSA

In this phase, we develop a wrapper based approach called RBGSA-MNB (recursive Binary Gravitational Search Algorithm-Multinomial Naive Bayes) to further refine the gene space obtained from the first phase.

Binary Gravitational Search Algorithm (BGSA) is a global optimal algorithm for dealing with discrete optimal problems [10,11]. BGSA has been successfully applied to gene selection in some real applications[12,13].

In BGSA, the location of the object i is represented through Eq.(3):

$$X_i = \left(x_i^1, \ldots, x_i^d, \ldots, x_i^s\right) \quad i = 1,2, \ldots, N \quad (3)$$

where N denotes the count of objects; x_i^d represents the ith object's location and s represents the count of dimensionality. Each object's mass is computed through Eq.(4):

$$M_i(t) = \frac{fit_i(t) - worst(t)}{\sum_{j=1}^{N}(fit_j(t) - worst(t))} \quad (4)$$

where $M_i(t)$ denotes the mass of object i; $f_i(t)$ is the fitness value of the object i; $worst(t)$ is the minimal value for maximization problems and is the maximal value for minimization problems:

$$worst(t) = \begin{cases} \min\limits_{j \in \{1,\ldots,N\}} fit_j(t) \ for \ maximization \ problems \\ \max\limits_{j \in \{1,\ldots,N\}} fit_j(t) \ for \ minimization \ broblems \end{cases} \quad (5)$$

The object's acceleration is computed through Eq. (6) and Eq. (7). Then object's next velocity is calculated by Eq. (8)). Finally, object's next location is computed through Eq. (9)

$$F_i^d(t) = \sum_{j \in kbest, j \neq i} rand_j G(t) \frac{M_j(t) M_i(t)}{R_{ij}(t) + \varepsilon} (x_j^d(t) - x_i^d(t)) \quad (6)$$

$$a_i^d(t) = \frac{F_i^d(t)}{M_i(t)} = \sum_{j \in kbest, j \neq i} rand_j G(t) \frac{M_j(t)}{R_{ij}(t) + \varepsilon} (x_j^d(t) - x_i^d(t)) \quad (7)$$

$$v_i^d(t+1) = rand_i \times v_i^d(t) + a_i^d(t) \quad (8)$$

$$x_i^d(t+1) = x_i^d(t) + v_i^d(t+1) \quad (9)$$

in which $rand_i$ and $rand_j$ denote two random number varying from 0 to 1; ε represents a positive constant; *kbest* includes K objects which have the largest mass.

In BGSA, $G(t)$ is a monotone descent function which is defined by Eq.(10):

$$G(t) = G_0\left(1 - \frac{t}{T}\right) \quad (10)$$

in which T represents the total counts of iteration. $R_{ij}(t)$ is the Euclidian distance between the object i and the object j which is calculated through Eq.(11):

$$R_{ij}(t) = \left\|X_i(t), X_j(t)\right\|_2. \quad (11)$$

The new location is computed through Eq.(12).

$$S\left(v_i^d(t)\right) = \left|tanh(v_i^d(t))\right| \quad (12)$$

After $S(v_i^d)$ is computed, agents will move according to Eq. (13).

$$\begin{aligned} &\textit{if } rand < S\left(v_i^d(t+1)\right) \\ &\quad \textit{then } x_i^d(t+1) = complement(x_i^d(t)) \\ &\textit{else } x_i^d(t+1) = x_i^d(t) \end{aligned} \quad (13)$$

As we mentioned before, gene selection using the wrapper based approaches can produce improved accuracy. However, these methods cost heavier cost of computation because they perform feature selection in a very large search space. Therefore, a good balance of exploration and exploitation is vital in the case of looking for a global optimal solution. In this work, we present a recursive BGSA for gene selection motivated by [18,19, 20,21]. This method first explores the whole feature set randomly, and then the set of features is optimized in the following steps and the BGSA is recursively performed on this reduced feature set while the accuracy is not declined.

The main idea of this process is to gradually transform a very raw gene set to a optimized one through removing gene at each step of the algorithm. The exploitation is enhanced through the recursive reduction of gene during the search. Algorithm 2 shows the recursive BGSA called R_BGSA. Algorithm 2 calls Algorithm 3 which is called R_BGSA_sub. In Algorithm 3, the standard BGSA is implemented in steps 1 -15. The stopping conditions are given in Steps 16 and 17. the gene set is cut down according to the global best agent in Step 19. In other words, those genes which are not chosen are eliminated from the gene set. In Step 20, the algorithm calls itself on the reduced gene set. Only when no more genes are cut down or accuracy begins declining, otherwise, the above process is done repeatedly. Finally, an optimized gene set and the global best accuracy are obtained. We use the Multinomial Naive Bayes to evaluate the fitness of the agents.

Algorithm 2: R_BGSA (D, s, N, ε, G_0, S , G_{acc}, *Flag*)
Input: Training gene set D, Number of genes s, Number of agents$N$, small constantε, gravitational constant G_0 **Output**: Set of selected genes S, Global best accuracy G_{acc}. 1. Initialize *Flag*=0. 2. Assign all genes to S, i.e. $\|S\|$= $s$and set global best accuracy G_{acc}= 0. 3. Call R_BGSA_sub (D, s, N,ε,G_0, S, G_{acc}, *Flag*) 4. Return selected gene subset S and global best accuracy G_{acc}
Algorithm 3: R_BGSA_sub (D, s, N, ε, G_0, S , G_{acc}, *Flag*)
Input: Training gene set D, Number of genes s, Number of particles N, small constant ε. gravitational constantG_0, Set of selected genes S, Global best accuracy achieved so far $G_{acc\text{-}so\text{-}far}$ **Output**: Set of selected genes S, Global best accuracy G_{acc} **1. for** each agent S_i **do** 2. Randomly produce a binary vector of length s for each agent in the population. 3. Construct dataset D for agent S_i (each agent denotes whether the feature is chosen or not), D consists of m samples (i= 1, 2, . . ., m), the features of each sample are selected based on S_i . 4. Divide dataset D into 10 disjoint sets D_v (v = 1, 2, . . ., 10) of similar size. 5. **for** each D_v do 6. Learn the classifier MNB_v on the basis of (D− D_v) 7. Calculate the accuracy of classification 8. **end for**

9. Evaluate fitness value (cross-validated accuracy)
10. Calculate the total acceleration a_i using Equation (8).
11. Calculate the velocity v_i through Equation (9).
12. Calculate the position x_i through Equation (10).
13. **end for**
14. Calculate *best* and $G_{acc\text{-}so\text{-}far}$.
15. **if** $G_{acc\text{-}so\text{-}far} \geq G_{acc}$, **then** S = *best* and $s' = |S|$.
16. **if** $|S| = s$, **then** *Flag*= 1.
17. **if** (*Flag*= 1 or $G_{acc\text{-}so\text{-}far} < G_{acc}$), **then** Stop.
18. Set $G_{acc} = G_{acc\text{-}so\text{-}far}$.
19. Decrease the gene set D to D' with chosen genes.
20. Call R_GSA_sele (D', s', N,ε, G_0, S, G_{acc}, *Flag*).

2.3. The ReliefF-RBGSA-MNB for cancer classification

Algorithm 4 is the ReliefF-RBGSA-MNB method.

Algorithm 4: ReliefF-RBGSA-MNB (D, s, N, ε, G_0, S , G_{acc}, *Flag*)

Input: Training gene set D, Number of genes m, Number of agents N, small constant ε, gravitational constant G_0
Output: Set of selected genes S, Global best accuracy G_{acc}.
1. Preprocess dataset by Eq.(3)
2. Select s genes using ReliefF approaches (Algorithm 1)
3. Decrease the gene set D to D' with chosen s genes.
4. Perform RBGSA on D' using fitness value, set the chosen genes to S and set the global best accuracy to G_{acc}. (Algorithm 2, Algorithm 3)
5. Obtain chosen gene subset S and global best accuracy G_{acc}.

3. Experimental Setup and Results

3.1. datasets and parameter settings

In this paper, we choose 6 microarray cancer datasets to evaluate our proposed ReliefF-RBGSA-MNB approach. Table 1 gives the simple description of cancer gene datasets, including the number of obtained samples, the number of genes of each sample and the number of classes. The parameters utilized in the RBGSA are shown in Table 2, which are chosen by performing many trials in order to get more objective value.

Table 1. Description of datasets.

Datasets	No. of total genes	No. of samples	No. of classes
Colon	2000	62	2
Central nervous system	7129	60	2
ALL-AML	7129	72	2
Breast	24,481	97	2
Lung	12,533	181	2
Ovarian	15,154	253	2

Table 2 Parameters for recursive Binary GSA algorithm.

Parameters	Values
Population size (P)	60
Total number of iterations (T)	100
small constant ε	3
gravitational constant G_0	100
v_{min}	−6
v_{max}	6

3.2. Results and discussion

In this section, we first analyze the effect of recursive feature elimination scheme by comparing ReliefF-RBGSA with ReliefF-BGSA on 6 microarray datasets. Then we evaluate the effectiveness of ReliefF-RBGSA-MNB by comparing our approach with 6 other techniques with regard to classification accuracy and number of genes selected on 6 microarray datasets.

3.2.1.comparison study with RBGSA and BGSA

In this subsection, We compare ReliefF-RBGSA with ReliefF-BGSA with regard to classification accuracy, average number of genes chosen, and Bonferroni correction P-values at 95% confidence to validate the effectiveness of recursive feature elimination scheme.

Figure. 1 show the results obtained using ReliefF-RBGSA and ReliefF-BGSA for 6 datasets. From these curves it can be seen that ReliefF-RBGSA obtained better performance in classification accuracy and got less number of selected genes.

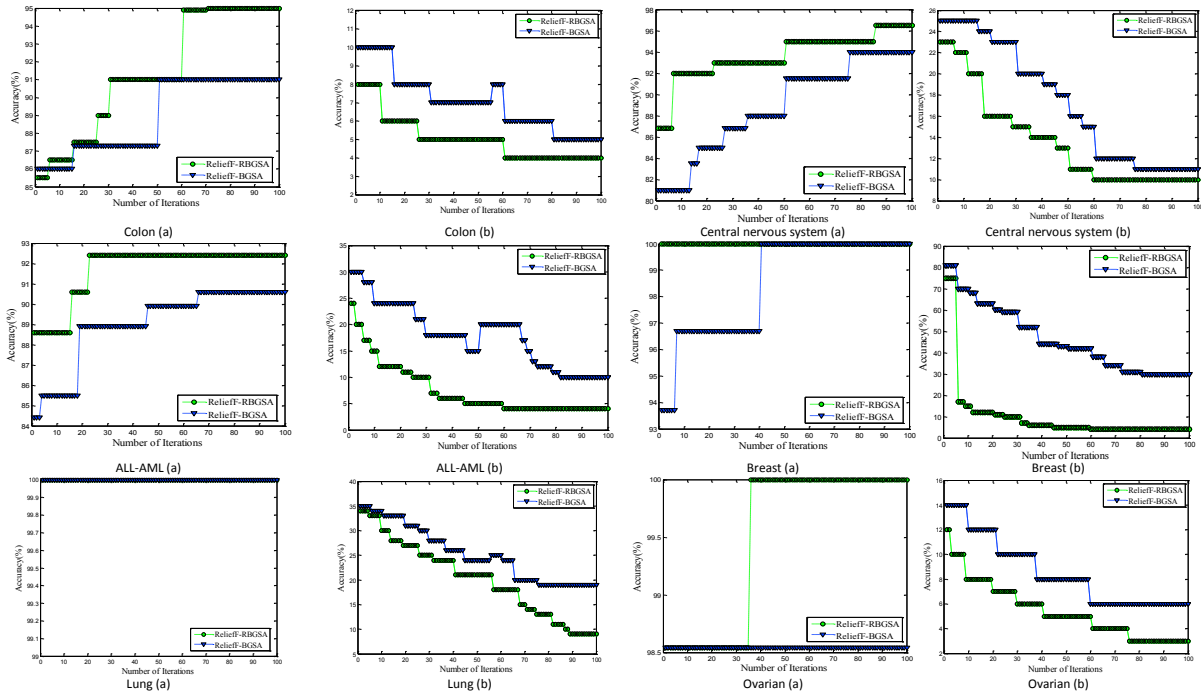

Fig. 1. Comparative evaluation curves of ReliefF-RBGSA and ReliefF-BGSA on classification accuracy (a) and number of genes (b).

Table 3 shows that the experimental results of classification accuracy and number of genes for the 6 microarray datasets got by ReliefF-RBGSA and ReliefF-BGSA. The classification accuracy is computed in terms of best value, standard deviation, and mean. In Table 3 Acc represents accuracy and NF indicates average number of genes selected in 10 runs. From Table 3, it is observed that the ReliefF-RBGSA achieve better performance in accuracy while choosing a very small set of genes in comparison to ReliefF-BGSA since it chooses the least number of relevant genes while not decreasing accuracy.

Table 3: Statistical results obtained by ReliefF-RBGSA and ReliefF-BGSA on 6 microarray datasets.

Datsets	Performance measures	ReliefF-RBGSA			ReliefF-BGSA		
		Best	Mean	SD	Best	Mean	SD
Colon	Acc	95.02	94.22	0.02	94.02	93.12	0.08
	NF	5	5.2	0.56	8	8.6	1.43
Central nervous system	Acc	100.00	100.00	0.05	99.00	98.89	0.07
	NF	9	9.7	0.89	11	11.4	0.92
ALL-AML	Acc	100.00	100.00	0.01	100.00	100.00	0
	NF	5	5.1	0.67	7	7.6	0.89
Breast	Acc	100.00	100.00	0.02	95.05	94.56	0.09
	NF	31	31.8	0.34	54	56.7	0.67
Lung	Acc	100.00	100.00	0.01	100.00	100.00	0
	NF	9	9.5	1.11	11	11.2	1.51
Ovarian	Acc	100.00	100.00	0	99.08	98.99	0
	NF	4	4.1	0.23	5	5.4	0.78

Table 4: P-values for ReliefF-RBGSA and ReliefF-BGSA for ten runs.

Datasets	ReliefF-RBGSA	ReliefF-BGSA
Colon	1.56e-07	4.56e-07
Central nervous system(CNS)	2.12e-08	1.45e-07
ALL-AML	6.43e-07	7.12e-06
Breast	4.78e-10	5.76e-10
Lung	4.12e-09	1.88e-08
Ovarian	1.98e-09	1.12e-08

We have presented statistical *p*-values for ReliefF-RBGSA and ReliefF-BGSA in Table 4. Frome Tables 4, we can see that ReliefF-RBGSA shows very small *p*-values over all datasets comparing to ReliefF-BGSA. From the p-values of Table 4 we can see that the accuracy difference for ReliefF-RBGSA method and ReliefF-BGSA is statistically significant.

3.2.2. comparison study between ReliefF-RBGSA-MNB and state-of-the-art Methods

In this subsection, we demonstrate the advance of our proposed ReliefF-RBGSA-MNB by comparing it with 6 other techniques including PSO-GA [22], HPSOTS [23], BMSF-SVM [24], RELIEF [25], FGM [26], FCLARANS [27].

We give the comparison results in Table 5 including the mean classification accuracy and average number of genes selected on 6 datasets. In table 5, CNS denotes Central nervous system. The ReliefF-RBGSA-MNB has obtained 98.39% in terms of classification accuracy among 6 approaches. The mean accuracy is computed on 10 independent run (each run includes 100 iterations) for each dataset individually. The highest mean classification accuracy is represented with bold type-face over each dataset. The ReliefF-RBGSA-MNB achieved 100% classification accuracy and have obtained comparative results on all datasets. The MBEGA method has the worst results in classification performance for each microarray cancer dataset.

Table 5 Experimental results of classification performance for 6 microarray cancer datasets.

Methods	Colon	CNS	ALL-AML	Breast	Lung	Ovarian	Average
PSO-GA	100	91.42	95.11	94.82	86.78	95.37	91.94
HPSOTS	93.7	91.92	92.86	82.07	90.29	91.74	91.04
BMSF-SVM	92.78	60.71	95.74	85.15	92.45	75.44	87.50
RELIEF	81.29	78.54	75.81	77.25	98.12	**100**	87.74
FGM	85.46	73.11	96.88	84.32	94.45	95.02	91.19
FCLARANS	**100**	56.53	82.03	89.19	92.67	95.65	86.57

4. Conclusion

In this work, we develop a novel method which integrates a ReliefF and RBGSA-MNB to perform classification of cancer. In our proposed method, the raw gene set is decreased first through ReliefF in order to remove the irrelevant and redundant genes. Then based on this decreased gene subset, RBGSA selects an optimized gene set using a recursive feature elimination scheme to obtain a great improvement of performance in terms of classification accuracy while cutting down a great deal of redundant features in the gene set, in which Multinomial Naive Bayes (MNB) classifier is used. We compared our proposed model with other existing methods for 6 publicly available benchmark datasets. Experimental results show that ReliefF-RBGSA-MNB achieves the best accuracy in most cases. Through reducing irrelevant and redundant genes, ReliefF-RBGSA-MNB effectively decreases the dimensionality of data. The obtained low dimensional set is the most important genes which can obtain higher classification accuracy.

Acknowledgments

This work was supported in part by the Natural Science Foundation of Shanxi Province (Grants No: 201801D121136) and nation Natural Science Foundation of China(61872260 and 61772358).

References

1. Perou C M, Jeffrey S S, Van d R M, et al. Distinctive gene expression patterns in human mammary epithelial cells and breast cancers[J]. Proceedings of the National Academy of Sciences of the United States of America, 1999, 96(16):9212-9217.
2. Golub T R, Slonim D K, Tamayo P, et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring, Science, 1999, 286 (5439) : 531–537.
3. Labani M, Moradi P, Ahmadizar F, et al. A novel multivariate filter method for feature selection in text classification problems[J]. Engineering Applications of Artificial Intelligence, 2018, 70:25-37.
4. Kira K, Rendell L A. A practical approach to feature selection[C]// International Workshop on Machine Learning. Morgan Kaufmann Publishers Inc. 1992:249-256.
5. Kononenko I, Estimating attributes: analysis and extension of relief, in: Proceedings of the Seventh European Conference in Machine Learning, Springer-Verlag, 1994: 171–182.
6. Robnik, Ikonja M, Kononenko I. Theoretical and Empirical Analysis of ReliefF and RReliefF[J]. Machine Learning, 2003, 53(1-2):23-69.
7. Kumar P G, Victoire T A A, Renukadevi P, et al. Design of fuzzy expert system for microarray data classification using a novel Genetic Swarm Algorithm[J]. Expert Systems with Applications, 2012, 39(2):1811-1821.
8. Nemati S, Basiri M E, Ghasem-Aghaee N, et al. A novel ACO–GA hybrid algorithm for feature selection in protein function prediction[J]. Expert Systems with Applications, 2009, 36(10):12086-12094.
9. Aghdam M H, Ghasem-Aghaee N, Basiri M E. Text feature selection using ant colony optimization[J]. Expert Systems with Applications, 2009, 36(3):6843-6853.
10. Rashedi E, Nezamabadi-Pour H, Saryazdi S. GSA: A Gravitational Search Algorithm[M]. Elsevier Science Inc. 2009.
11. Rashedi E, Nezamabadi-Pour H, Saryazdi S. BGSA: binary gravitational search algorithm[J]. Natural Computing, 2010, 9(3):727-745.
12. Nagpal S, Arora S, Dey S, et al. Feature Selection using Gravitational Search Algorithm for Biomedical Data[J]. Procedia Computer Science, 2017, 115:258-265.
13. Han X H, Chang X M, Quan L, et al. Feature subset selection by gravitational search algorithm optimization[J]. Information Sciences, 2014, 281:128-146.
14. Xiang J, Han X H, Duan F, et al. A novel hybrid system for feature selection based on an improved gravitational search algorithm and k-NN method[J]. Applied Soft Computing, 2015, 31(C):293-307.
15. Bababdani B M, Mousavi M. Gravitational search algorithm: A new feature selection method for QSAR study of anticancer potency of imidazo[4,5-b]pyridine derivatives[J]. Chemometrics & Intelligent Laboratory Systems, 2013, 122(5):1-11.
16. Baranovsky A, Daems D. DESIGN OF ONE-DIMENSIONAL CHAOTIC MAPS WITH PRESCRIBED STATISTICAL PROPERTIES[J]. International Journal of Bifurcation & Chaos, 1995, 5(06):1585-1598.
17. Kohavi R, John G H. Wrappers for feature subset selection[J]. Artificial Intelligence, 1997, 97(1-2):273-324.
18. Guyon I, Weston J, Barnhill S, et al. Gene Selection for Cancer Classification using Support Vector Machines[J]. Machine Learning, 2002, 46(1-3):389-422.
19. Li F, Yang Y. Analysis of recursive gene selection approaches from microarray data[J]. Bioinformatics, 2005, 21(19):3741-7.
20. Mundra P A, Rajapakse J C. SVM-RFE with MRMR filter for gene selection.[J]. IEEE Transactions on Nanobioscience, 2010, 9(1):31-37.
21. Prasad Y, Biswas K K, Hanmandlu M. A Recursive PSO scheme for Gene Selection in Microarray Data[J]. Applied Soft Computing, 2018.
22. Li S, Wu X, Tan M. Gene selection using hybrid particle swarm optimization and genetic algorithm[J]. Soft Computing, 2008, 12(11):1039-1048.

23. Shen Q, Shi W M, Kong W. Hybrid particle swarm optimization and tabu search approach for selecting genes for tumor classification using gene expression data[J]. Computational Biology & Chemistry, 2008, 32(1):53-60.
24. Zhang H, Wang H, Dai Z, et al. Improving accuracy for cancer classification with a new algorithm for genes selection[J]. BMC Bioinformatics, 2012, 13(1):1-20.
25. Kira K, Rendell L A. The feature selection problem: traditional methods and a new algorithm[C]// Tenth National Conference on Artificial Intelligence. AAAI Press, 1992:129-134.
26.Tan M, Wang L, Tsang I W. Learning sparse SVM for feature selection on very high dimensional datasets[C]// International Conference on International Conference on Machine Learning. Omnipress, 2010:1047-1054.
27. Ghosh S, Mitra S, Dattagupta R. Fuzzy clustering with biological knowledge for gene selection[M]. Elsevier Science Publishers B. V. 2014.