Biochemical and Biophysical Research Communications 446 (2014) 850–856

Contents lists available at ScienceDirect

Biochemical and Biophysical Research Communications

journal homepage: www.elsevier.com/locate/ybbrc

Robust and stable feature selection by integrating ranking methods and wrapper technique in genetic data classification

Maryam Yassi [a,*], Mohammad Hossein Moattar [b]

[a] Young Researchers and Elite Club, Mashhad Branch, Islamic Azad University,, Mashhad, Iran
[b] Department of Software Engineering, Mashhad Branch, Islamic Azad University, Mashhad, Iran

ARTICLE INFO

Article history:
Received 19 February 2014
Available online 19 March 2014

Keywords:
Microarray classification
Dimension reduction
Filter method
Wrapper method
Support vector machine
Imbalance classes

ABSTRACT

High dimensional data increase the dimension of space and consequently the computational complexity and result in lower generalization. From these types of classification problems microarray data classification can be mentioned. Microarrays contain genetic and biological data which can be used to diagnose diseases including various types of cancers and tumors. Having intractable dimensions, dimension reduction process is necessary on these data. The main goal of this paper is to provide a method for dimension reduction and classification of genetic data sets. The proposed approach includes different stages. In the first stage, several feature ranking methods are fused for enhancing the robustness and stability of feature selection process. Wrapper method is combined with the proposed hybrid ranking method to embed the interaction between genes. Afterwards, the classification process is applied using support vector machine. Before feeding the data to the SVM classifier the problem of imbalance classes of data in the training phase should be overcame. The experimental results of the proposed approach on five microarray databases show that the robustness metric of the feature selection process is in the interval of [0.70, 0.88]. Also the classification accuracy is in the range of [91%, 96%].

1. Introduction

Applying microarray technology which makes studying the expression of thousands of genes possible simultaneously, has led to the production of massive amounts of gene expression data recently. Statistical analysis of these data includes feature selection, normalization, and classification.

The expression levels of genes have important information about biological networks, cellular states, and the understanding of gene function. An objective of gene expression data analysis is to determine how the expression of each individual gene affects the expression of other genes or genetic networks. Another goal is to determine how these genes are expressed in healthy and diseased cells. Using data mining techniques and artificial intelligence techniques for analyzing data obtained from this technology can be useful for diagnosis and treatment.

Collections of genetic data have high dimensions and small size and are usually imbalanced. Data collection and analysis and discovering unknown relations among these data are complex tasks. High dimensions increase the complexity of the classification process and the prediction of disease type and since there are redundancies and duplications of genes, the classification accuracy may degrade. Experiments which are carried out to extract the gene expression matrix are costly and due to the limited number of experiments, we are faced with the small size of the data set. The small size of the data set leads to another challenge. Among those people that are tested most are not likely to have the disease and those which are suspicious to have cancer or tumor are less likely. In this case, we are faced with minority in genetic databases who are sick.

Different strategies have been proposed over the last years for feature selection, including filter, wrapper [28], embedded [29], and more recently ensemble techniques [30].

The feature selection is an essential problem for pattern classification. The proposed feature selection approach is based on ranking methods [2] to select each characteristic individually, and without considering the relationship between features while other feature selection approaches are based on the selection of the best subset of features considering the interaction between the features [3].

Feature ranking based selection methods evaluate the significance of features according to some measurements, such as distance [31,32], and information theory [33]. Among the distance based measures, Relief, which is firstly proposed by Kira and

* Corresponding author.
E-mail addresses: Yassimaryam@gmail.com (M. Yassi), Moattar@mshdiau.ac.ir (M.H. Moattar).

http://dx.doi.org/10.1016/j.bbrc.2014.02.146
0006-291X/

Rendell [31] is one of the most successful ones and adopt Euclidean distance to assign a relevance weight to each feature. The key idea of Relief is to iteratively estimate feature weights according to their ability to discriminate between nearby instances. However the optimality of Relief is not guaranteed because Relief randomly picks out an instance from training dataset.

Subset selection algorithms search the set of possible features for the optimal subset in which features are relevant in the given model. One critical problem for feature subset selection methods is that exhaustive search and evaluation of all the possible feature subsets, which usually ends in a considerably high computational complexity [34]. Thus, many heuristic subset search strategies have been introduced [35], such as sequential forward/backward selection, random selection [36], and branch and bound search [37]. A good feature subset is one that contains features highly correlated with the class, yet uncorrelated with each other [38].

Robust feature selection algorithms are related to the sensitivity of the algorithm facing with various real world conditions, such as disruptions in the training data. If a feature selection algorithm is not robust, increasing and decreasing the training samples will lead to different results. To enhance the strength of feature selection and not to get stuck in local optima, authors of [4,5] describe the integration of various criteria to build a more robust and appropriate dataset. The point which is an important factor in the feature selection algorithm is the stability of feature selection.

Considering imbalanced databases is a comprehensive look at the real world. Authors of [6] offer a feature ranking method based on density estimation to deal with the problem of imbalanced classes. The authors of [7], using genetic algorithm combined with fuzzy theory to select the most distinctive features. The Huang-index method using fuzzy c-means is employed to enhance cluster validity and achieve consistent clusters of the features. Also a new entropy-based feature evaluation method is formulated for the authentication of relevant features. Then, multivariate statistical analyses are utilized to solve the redundancy between relevant features [39]. Criteria to describe the relationship between features includes: Entropy, Mutual information and Information-F.

The authors of [8] determine the relevance of each feature with other features using similarity relationships among the data sets. The method used is based on an unsupervised process. First, the discrimination of each feature is calculated. In the next, step of the features that are rated higher in terms of discrimination are selected in an iterative process. Authors of [40] offer a dynamic weighting-based feature selection algorithm, which not only selects the most relevant features and eliminates redundant features, but also tries to retain useful intrinsic groups of interdependent features. The primary characteristic of the method is that the feature is weighted according to its interaction with the selected features and the weight of features will be dynamically updated after each candidate feature has been selected. The authors of [9,10] try to overcome the weaknesses of theoretical methods using cooperative game. This means that if the discrimination of a single feature is weak, it will improve when placed alongside other features. Therefore, this method represents the strength of features to make any discrimination imbalanced nature of the dataset is another problem in this field.

Due to the great importance that today genetic data has, it is are required to present a method which has an appropriate performance and also can overcome mentioned challenges. The method proposed in this paper tries to solve the problems of complexity and overcome the above mentioned challenges. One of the challenges related to high dimensional data sets of genetic is the dimension reduction based on feature selection to specify distinct and superior genes.

The first stage of the feature selection process consists of two parts. In the first part, the process of feature selection integrates ranking methods. The integration of ranking methods causes better robustness and feature selection stability. In the second part of the feature selection process, a wrapper technique is applied [1] which has the ability to express interactions between genes. In the first part, the unique characteristics of selected features are analyzed; while in the second part the interaction between features are analyzed.

The second phase of the proposed approach is based on SVM classifier. Using SVM classifier is due to the high generalization ability. As previously mentioned, genetic data sets have small sizes and low generalization ability. Thus, SVM is appropriate to be used to overcome this problem. As previously mentioned, the genetic datasets are imbalanced and SVM is highly sensitive against imbalanced datasets. Separating hyper plane would not be located properly between data of two classes and will be close to the majority class. In this paper, a method is proposed to solve this problem by removing the data points from majority class which are far from the decision boundary.

Organization of the materials presented in this paper is as follows: In the Section 2 the proposed method is introduced. In Section 3, the simulation results obtained from the proposed method are discussed. Conclusions and future works are described in Section 4.

2. Materials and methods

Due to the complexity of high-dimensional problems, we need to provide an intelligent way to select appropriate features. The best feature set is that with the highest performance and the lowest classification error.

2.1. Ranking method

The proposed method of feature selection is grounded on filtering techniques based on the ranking of features. Top ranked features which are selected based on statistical approaches and represent enough information and functionality to enhance the performance of classification is chosen. Ranking methods used in this study is tabulated in Table 1.

In the above mentioned, x_i is attribute values of ith sample, $\bar{x}_1$ is its average value and σ_{x_i} the characterizes the variance of samples. C is the class label and parameters such as n_1 and n_2 demonstrates the number of samples belonging to any particular features in the corresponding class label. Also S_W and S_B are the within-class and between-class scatter matrix, respectively.

2.2. Integrating ranking methods

The voting system we are going to use in this study merges the output of each of the ranking methods. The output of each method is a ranking list in the descending order. In this voting system, voters are ranking functions and volunteers are the entire set of features. Finally, the integration of the outputs of the ranking methods is a list of features sorted by the votes earned from each ranking function.

Our proposed voting system determines which features are in the top of the ranking functions. Distinctive and more superior features are selected based on total votes received for that feature. This study develops an integration method which provides the stability of feature selection process.

2.3. Integrating wrapper method with ranking methods

Integrating ranking methods using voting system is already investigated. But feature relations and response characteristics has not been studied. To combine the integrated ranking method

Table 1
Ranking methods.

Refs.	Criterion	Name
[11]	$S = \mathrm{Info}(X) - \mathrm{Info}_x(X)$ $\mathrm{Info}(X) = -\sum_{i=1}^{k} P(c_i, X) \times log(P(c_i, X))$ $\mathrm{Info}(X) = -\sum_{i=1}^{V} \frac{\lvert V_i \rvert}{\lvert X \rvert} \times \mathrm{Info}(V_i)$ K – number of classes, v – number of individual values of a genes x V_i – the set of instances whose values in gene x equal x_i	Information gain
[12]	$\lvert V_i \rvert$ – number of samples in $\lvert V_i \rvert, \lvert X \rvert = n$ $S_w = \sum_{i=1}^{c}\sum_{j}^{n_c}(x_j - \mu_i) * ((x_j - \mu_i))^T$ $S_b = \sum_{i=1}^{c}(\mu_i - \mu) * ((\mu_i - \mu))^T$ $S = \frac{S_B}{S_W}$	Between versus within class scatter ratio
[13]	$\mathrm{FR}(x) = \frac{([\bar{x}_1 - \bar{x}_2])^2}{\sigma^2_{\bar{x}_1} + \sigma^2_{\bar{x}_2}}$	Fisher ratio
[14]	$S = \frac{\bar{x}_1 - \bar{x}_2}{\sigma_{x_1} + \sigma_{x_2}}$	*T*-test
[15]	$S = \frac{\bar{x}_1 - \bar{x}_2}{\sigma_x}$	Z-score
[16]	$S = \bar{x}_1 - \bar{x}_2$	Fold-change difference
[17]	$S = \frac{\bar{x}_1}{\bar{x}_2}$	Fold-change ratio
[18]	$S = \sum_{j=1}^{k} R_j, K = \min(n_1, n_2)$	Wilconxon rank sum
[19]	$S = \left(\prod_{j=1}^{n} R_j\right)^{\frac{1}{n}}$	Rank product
[20]	$S = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\sigma_{x_1}/n_1 + \sigma_{x_2}/n_2}}$	Welch-*t*-test

with wrapper method, feature selection is done in an iterative process based on an evaluation criterion for selecting a subset of superior features. After integrated ranking method, a large number of attributes will be eliminated due to fewer votes. However, by eliminating many of the features, iterative wrapper process will benefit greater speed. The purpose of this section is to consider the interactions between the remaining features. Selected features of the integrated method, is sorted in the descending order based on the votes taken in the previous step. Based on the ordered list, the first distinctive feature is initially added to an empty set of features. To add an attribute to a subset of distinctive features from the top ranked features, between class scatter to within-class scatter ratio is calculate according to Eq. (2). Taking the mentioned measure into account the feature which increases the measure is added to the feature set. The process is repeated until the determined number of features is selected. In addition to reducing the size of feature space, this processes results in the selection of subsets of the distinguishing features of the strength and stability.

2.4. Support vector machine classifier

Support vector machine is generally used for issues in which there are two classes. Of course, different ways of combining multiple SVM classifiers have been proposed. SVM algorithm locates a linear boundary between two classes of data and then solves the problem of finding the maximum margin between these two data sets. Basically, one of the strengths of SVM is the nonlinear mapping of the input vectors into a high dimensional feature space, in which the points are linear separable. This is done by kernel functions. The Gaussian kernel function used for SVM classifier in this study is introduced in Eq. (1). The reason of using this type of kernel function is because of the flexibility that enables better separation than other kernel functions.

$$K(u, v) = \left(-\frac{1}{2\gamma^2}(u - v)\right) \quad (1)$$

2.5. Imbalanced data classification

When the classes are imbalanced the majority of data belongs to one of the classes. SVM classifier is sensitive to the imbalanced data. Boundary plates are not considered if the number of instances of the class is small. This leads to the increment of the error rate. In this condition, detection of minority class will be incorrect. To solve this problem, various methods have been proposed [21,22], An approach is to add samples to data points of the minority class. The method which is proposed in this paper to solve the problem of imbalanced classes based on the removal of the data points of the majority class. Samples of the majority class which are far from the decision boundary are removed. To determine which data in the majority class is far from decision boundary, each time, for all samples in the majority class, one sample is omitted and considering the removed data, between class scatter matrixes (S_B) for both the minority and majority classes are calculated. Finally, the sample creating the greatest S_B is deleted from the data set of the majority class. This process continues until there is a balance between the numbers of data points of the minority class and the majority class.

3. Results and discussion

3.1. Data set

In the experimental study, extensive experiments were conducted on the following five gene-expression data sets:

Prostate data [23]: This data set contains expression level of 12,600 genes for 136 samples including 77 prostate tumors, and 59 normal samples.

DLBCL data [24]: This set of data contains 58 DLBCL (diffuse large b-cell lymphoma) samples and 19 FL (Follicular Lymphoma) samples with 7,129 genes.

Colon data [25]: This data set contains expression levels of 2000 genes for 62 samples including 22 normal samples, and 40 colon cancer samples. The task is to distinguish between normal and tumor samples.

Leukemia data [26]: Introduced by Golub et al., in 1999, this data set contains expression levels of 7129 genes for 47 ALL (Acute lymphoblastic leukemia) leukemia patients and 25 AML (Acute myelogenous leukemia) leukemia patients.

CNS data [27]: The goal of this study is the molecular investigation of treatment effectiveness for embryonal CNS (Central Ner-

Table 2
Data sets characteristics.

Name	# Class 1	# Class 2	# Feature
Prostate	59	77	12,600
DLBCL	19	58	7129
Colon	22	40	2000
Leukemia	25	47	7129
CNS	21	39	7129

Table 3
Confusion matrix.

Name	Predicted positives	Predicted negatives
Peal positives	TP	FN
Peal negatives	FP	TN

vous System) tumors. The task is to distinguish between failed and succeed treatment outcomes. There are 60 patients with 7129 genes in this data set, where 21 patients are survivors and 39 patients are failures. These five data sets are briefly summarized in Table 2.

Table 4
Avearge accuracy of KNN classifier on genetic data sets.

Number of feature	Info gain	B_W class scatter ratio	Fisher ratio	*T*-test	*Z*-score	Fold change difference	Fold change ratio	Wilcoxon rank sum	Rank product	Welch *t*-test	Fusion
5	77.25284	79.21362	61.14642	71.84184	67.36932	68.72518	68.3392	80.02704	60.2039	70.8427	**80.90802**
10	72.9397	73.72402	68.8766	69.92916	66.28894	71.09794	72.6984	77.79054	61.99812	69.72814	**80.52424**
15	71.61758	71.61758	63.67978	66.08072	71.49816	65.984	67.19318	77.34798	61.2486	69.00014	**81.5934**
20	68.41338	67.23692	61.41558	64.13112	65.50968	65.27632	68.78862	73.9235	61.18198	68.71944	**78.92672**
25	67.37726	68.55372	62.60356	68.65714	65.25372	65.11446	68.64328	72.08038	60.04944	67.98762	**78.4004**
30	65.58924	65.19708	62.55344	66.14732	60.44168	64.60464	66.52152	69.02894	58.21868	62.58054	**78.01946**
35	63.11304	62.72088	59.4433	64.98178	60.2061	63.80382	65.19228	67.03514	57.43318	62.14946	**82.44342**
40	62.83146	62.83146	61.17908	66.95818	60.77754	63.64106	60.35784	68.46462	54.952	62.91138	**82.69524**
45	60.61856	61.40288	61.08384	65.41194	66.31784	61.28162	60.40298	66.24052	55.87044	60.6811	**79.73256**
50	60.14238	59.35808	61.08384	66.07862	59.32034	65.7475	60.93546	64.38086	57.20378	60.01442	**79.73256**

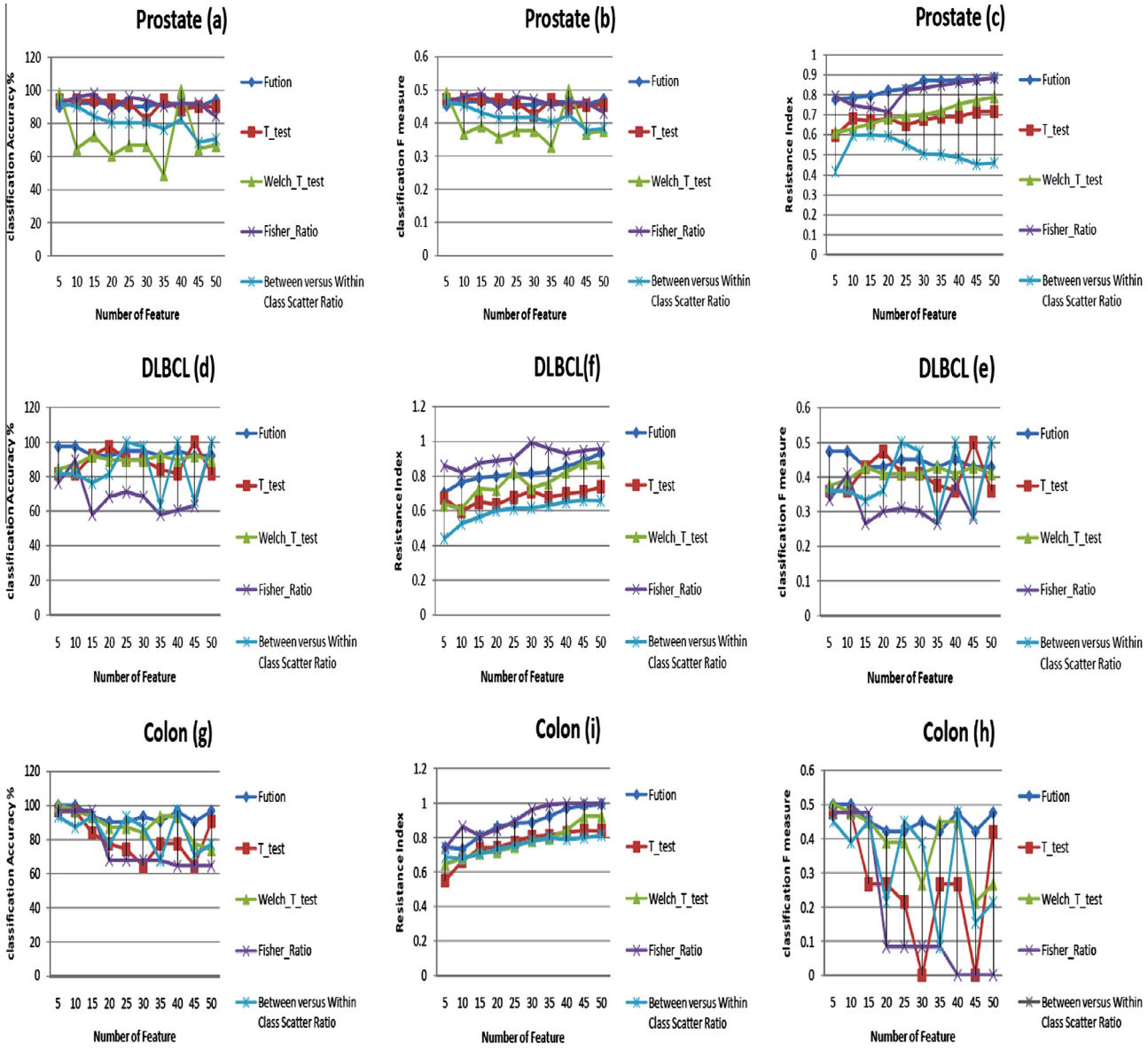

Fig. 1. Accuracy, *F* measure and Robustness index on genetic data sets using SVM classifier.

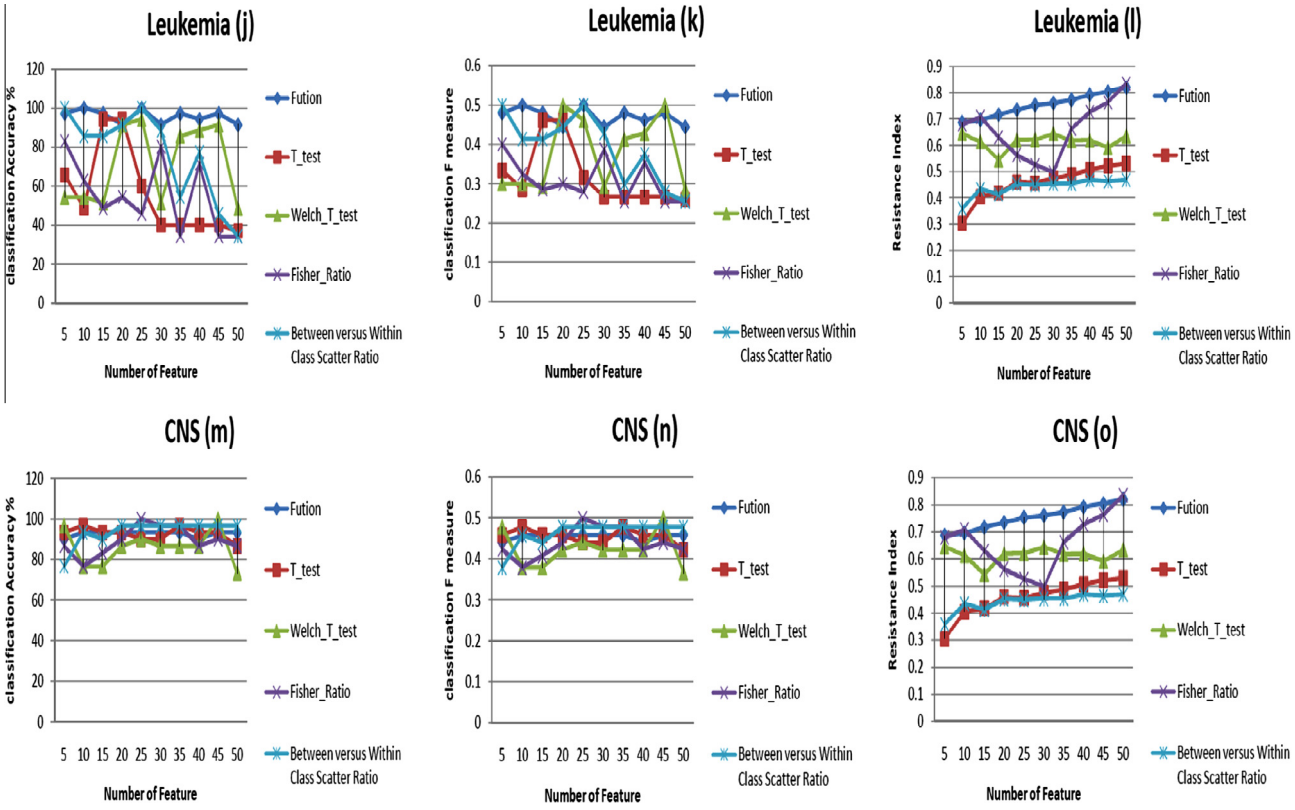

Fig. 1 (*continued*)

3.2. Evaluation criteria

Criteria used to evaluate the proposed method are stability for feature selection process, classification accuracy and *F* measure.

3.2.1. Accuracy of classification

Classification accuracy is the main criterion for evaluating the classification and prediction of samples in the test phase. Table 3 defines the main measurement for accuracy when the samples are correctly or incorrectly classified. In this investigation the majority class is considered as the negative class and minority class is the positive class. The performance of the two-class classification problems is calculated according to Eq. (2).

$$\text{accuracy} = (\text{TP} + \text{TN})/(\text{TP} + \text{TN} + \text{FN} + \text{FP}) \tag{2}$$

3.2.2. F measure

On the other hand, sometimes we are interested in highly effective detection ability for only one class. For such problems, another pair of metrics, precision as Eq. (3) and recall as Eq. (4), is often adopted. Notice that recall is the same as sensitivity. *F*-Measure as Eq. (5) is used to integrate precision and recall into a single metric for convenience of comparison. The parameter β is introduced as a regulatory factor and its value is equal to one in this research.

$$\text{recall} = \frac{\text{TN}}{\text{TN} + \text{FN}} \tag{3}$$

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{4}$$

$$F\text{-measure} = \frac{(1 + \beta^2) \times \text{recall} \times \text{precision}}{\beta^2 \times \text{recall} + \text{precision}} \tag{5}$$

3.2.3. Robustness measure

Measuring feature selection robustness would be beneficial. Assume that *m* sets of data with different repetitions of each data sample and a subset of the distinguishing features is selected. For measuring more accurate overall similarity between each set with the others, Eq. (6) is use. S_i and S_j parameters are two subsets of selected characteristics.

$$JC_i(k) = \frac{|S_i \cap S_0| + SC_i}{k} \tag{6}$$

SC_i is equal to the sum of the absolute correlation between selected features. $|S_i \cap S_0|$ is equal to the number of common features between the two selected set of features and *K* is equal to set size of S_i and S_0. Assume totally m batches of data are generated by resembling and m feature subsets are selected. Now, robustness of the feature selection algorithm is calculated by Eq. (7). The robustness index is in the interval of [0,1].

$$\overline{JC}(k) = \frac{\sum_{i=1}^{m} JC_i(k)}{m} \tag{7}$$

3.3. Evaluation of feature selection criteria

Based on the average accuracy of the classification performed on the least number of distinct genes the Fisher method, the between versus within class scatter ratio, the Welch-test and *T*-test methods have comparable results with the proposed approach. Hence in the next parts, we will compare the proposed method only against the four mentioned measures.

3.3.1. KNN classifier

In the first experiment the 5-nearest neighbor classifier is applied on the microarray datasets. Averaging the results of the experiments of the ranking methods based on genetic databases

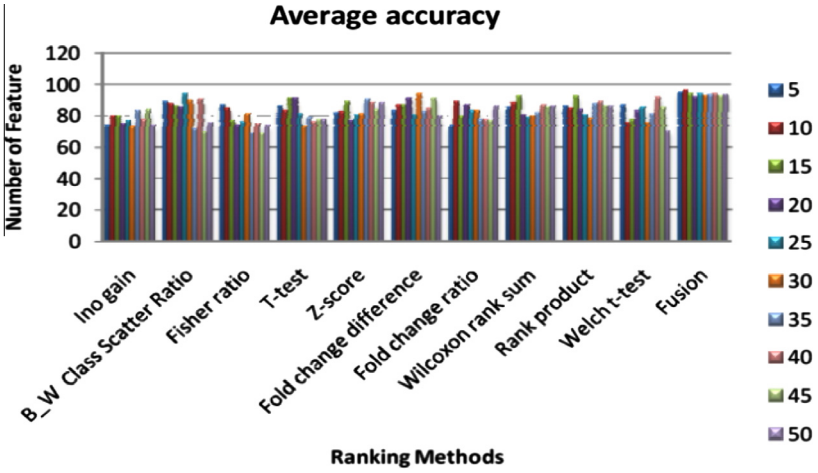

Fig. 2. Average accuracy of the proposed approach using SVM classifier on genetic data set.

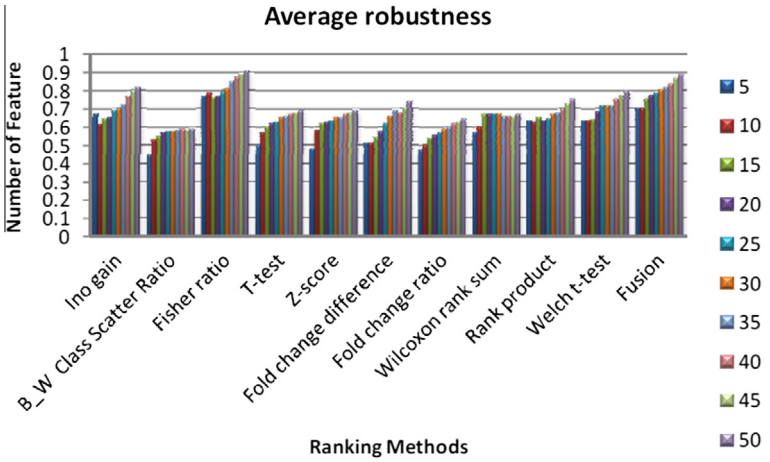

Fig. 3. Average robustness of the proposed approach using SVM classifier on genetic data set.

using KNN classifier is illustrated in Table 4. It shows that the proposed feature selection process is more superior and robust compared to the other approaches regardless of the type of classifier. Using the classifiers other than SVM shows that the proposed method in this paper is independent of the type of classifier and the proposed process of feature selection is more robust and stable than the previously proposed methods.

3.3.2. SVM classifier

In the second experiments, the proposed method is evaluated on five microarray datasets using SVM classifier and considering all the previously mentioned performance measures including classification accuracy, *F* measure and Robustness of the feature selection process.

In (Fig. 1 – Parts A,B,G,H), as the number of features increases, stability and robustness become observable in classification accuracy and *F* measure on the prostate and colon data sets. The absence of such robustness in other methods compared with the proposed method is also observable. In (Fig. 1 – Parts D,E,J,K), classification accuracy and *F* measure criteria of the integrated ranking method on DLBCL and Leukemia data sets, change slowly as the number of selected features change, while changes are very rapid and sudden with low robustness for the other methods. However, in Fig. 1 – Parts M, N, the proposed method has a moderate performance compared with other methods. Its stability along with the slow changes in the performance criteria is outstanding.

In Fig. 1 – Parts C, L, O, the stability of the proposed method indicates its better performance in the training phase when there is perturbations of training data. In Fig. 1 – Parts I, F, the robustness of the proposed method is lower compared with the Fisher measure. However, Fisher criterion has a poor performance on DLBCL and colon data sets in the other two criteria. The purpose of the proposed method is improving the classification accuracy and *F* measure criteria while preserving the stability of feature selection process.

Average accurately of the SVM classifier on the 5 Databases is shown in Fig. 2. As shown, for each number of selected features, the proposed method has higher average accuracy compared to other methods and also shows less variation among the other methods which suggests its better robustness.

The robustness index as depicted in Fig. 3 is comparable with the method of Fischer rate; however, higher robustness of the Fisher method cannot guarantee its higher accuracy. Having these observations the proposed method has acceptable strength and is favorable in the feature selection process and also has high classification accuracy.

As a summary, the robustness index of the feature selection process of the proposed approach is in the interval [0.70,0.88] and the classification accuracy is in the range of 91–96%. Limited range of changes of the mentioned criteria will prove the sustainability of the proposed method.

4. Conclusion

In this paper, a method for reducing dimensions of high-dimensional data sets is presented. Feature selection process has been done based on the integration of ranking and wrapper techniques. The proposed method is independent of the type of classifier and possesses desirable stability for the change in the number of selected features. Due to the high flexibility and lack of sensitivity to data set with small size, SVM is proposed for classification. Before feeding of genetic datasets to SVM classifier, the problem of imbalanced classes is solved with the removal of the majority class samples that are far from the decision boundary. The average results of the proposed method on five microarray databases realizes that robustness index of the feature selection process is in the interval of [0.70, 0.88], the classification accuracy is in the interval [91%, 96%] and F measure is in the range [0.44, 0.47] for every change in the number of selected features. Indeed, the limited range of criteria represents the stability of the proposed method. As the future works, it may be worth studying a more than two-class problem and combining the SVM classifier results using methods such as comparative games in Game theory [41]. Also using the embedded techniques such as decision tree for extracting interactions between genes may be beneficial.

Acknowledgments

We would like to thank Dr. Mehdi Yaghoobi, Islamic Azad University, Mashhad Branch, for helpful discussions and critical reading of the manuscript.

References

[1] R. Jensen, Q. Shen, Computational Intelligence and Feature Selection, IEEE Press Series On Computational Intelligence, Canada, 2008.
[2] J.T. Cosmin lazar, A survey on filter techniques for feature selection in gene expression microarray analysis, IEEE/ACM Trans. Comput. Biol. Bioinf. 14 (2012).
[3] R. Ruiz, J.C. Riquelme, J.S. Aguilar-Ruize, M. Garcia-Torres, Fast feature selection aimed at high-dimensional data via hybrid-sequential-ranked searches, Expert Syst. Appl. 9 (2012).
[4] F.Y. Mao, Robust feature selection for microarray data based on multicriterion fusio, IEEE/ACM Trans. Comput. Biol. Bioinf. 13 (2012).
[5] T.M. Waelawada, A Review of The Stability of Feature Selection Techniques for Bioinformatics Data, Las Vegas, Nevada, USA, IEEE, 2012. p. 8.
[6] S.H. Minaalibeigi, DBFS: an effective Density Based Feature Selection scheme for small sample size and high dimensional imbalanced data sets, Data Knowledge Eng. 37 (2013).
[7] M.I. Gerald Schaefer, Data mining of gene expression data by fuzzy and hybrid fuzzy methods, IEEE Trans. Inf. Technol. Biomed. 7 (2010).
[8] Artur J. Ferreira, Mário A.T. Figueiredo, Efficient feature selection filters for high-dimensional data, Pattern Recognit. 11 (2012).
[9] Y.L. Xin Sun, Feature evaluation and selection with cooperative game theory, Pattern Recognit. Lett. 11 (2012).
[10] Y.L. Xin Sun, Using cooperative game theory to optimize the feature selection problem, Neurocomputing 8 (2011).
[11] L.-Y. Chuang et al., A two-stage feature selection method for gene expression data, OMICS: J. Integr. Biol. 13 (2009) 127–137.
[12] R. Fisher, The use of multiple measurements in taxonomic problems, Ann. Eugen. 7 (1939) 179–188.
[13] T.X. JiepingYe, Computational and theoretical analysis of null space and orthogonal linear discriminant analysis, J. Mach. Learn. Res. 22 (2006).
[14] J.G. Thomas et al., An efficient and robust statistical modeling approach to discover differentially expressed genes using genomic expression profiles, Genome Res. 11 (7) (2001) 1227–1236.
[15] T.R. Golub et al., Molecular classification of cancer: class discovery and class prediction by gene expression monitoring, Science 286 (5439) (1999) 531–537.
[16] M.K. Kerr, M. Martin, G.A. Churchill, Analysis of variance for gene expression microarray data, J. Comput. Biol. 7 (6) (2000) 819–837.
[17] H. Tao et al., Functional genomics: expression analysis of *Escherichia coli* growing on minimal and rich media, J. Bacteriol. 181 (1999) 6425–6440.
[18] L. Deng et al., A rank sum test method for informative gene discovery, Proc. 10th ACM SIGKDD Int'l Conf. Knowl. Disc. Data Min. (2004) 410–419.
[19] R. Breitling et al., Rank products: a simple, yet powerful, new method to detect differentially regulated genes in replicated microarray experiments, FEBS Lett. 573 (1–3) (2004) 83–92.
[20] S. Dudoi et al., Statistical methods for identifying differentially expressed genes in replicated cDNA microarray experiments, Stat. Sinica 12 (2002) 111–139.
[21] C.-W. Chih-FongTsain, SVOIS: Support Vector Oriented Instance Selection for text classification, Inf. Syst. 14 (2013).
[22] D.G. Zeping Yang, An active under-sampling approach for imbalanced data classification, in: Fifth International Symposium on Computational Intelligence and Design, IEEE, 2012 (p. 4).
[23] D. Singh et al., Gene expression correlates of clinical prostate cancer behavior, Cancer Cell 1 (2) (2002) 203–209.
[24] M.A. Shipp et al., Diffuse large b-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning, Nat. Med. 8 (1) (2002) 68–74.
[25] U. Alon, N. Barkai, D.A. Notterman, K. Gishdagger, S. Ybarradagger, D. Mackdagger, A.J. Levine, Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays, Proc. Natl. Acad. Sci. USA 96 (12) (June 1999) 6745–6750.
[26] T. Golub et al., Molecular classification of cancer: class discovery and class prediction by gene expression monitoring, Science 286 (5439) (Oct. 1999) 531–537.
[27] S.L. Pomeroy et al., Prediction of central nervous system embryonal tumor outcome based on gene expression, Nature 415 (2002) 265–271.
[28] R. Jensen, Q. Shen, Computational Intelligence And Feature Selection, IEEE Press Series On Computational Intelligence, Canada, 2008.
[29] Y. Saeys, I. Inza, P. Larrañaga, A review of feature selection techniques in bioinformatics, Bioinformatics 23 (19) (2007) 2507–2517.
[30] P. Yang et al., A review of ensemble methods in bioinformatics, Curr. Bioinfor. 5 (4) (2010) 296–308.
[31] K. Kira, L.A. Rendell, A practical approach to feature selection, in: Proceedings of the Ninth International Workshop on Machine Learning, Morgan Kaufmann Publishers Inc., Scotland, United Kingdom, 1992, pp. 249–256.
[32] S.W. Card, , Information distance based fitness and diversity metrics, in: Proceedings of the 12th Annual Conference Companion on Genetic and Evolutionary Computation, ACM, Portland, Oregon, USA, 2010, pp. 1851–1854.
[33] F.E. BoyanBonev, Information-theoretic selection of high-dimensional spectral features for structural recognition, Comput. Vis. Image Und. 15 (2012).
[34] S. Davies, S. Russell, NP-completeness of searches for smallest possible featuresets, in: AAI Symposium on Intelligent Relevance, AAAI Press, New Orleans, 1994, pp. 41–43.
[35] H. Liu, L. Yu, Toward integrating feature selection algorithms for classification and clustering, IEEE Trans. Knowl. Data Eng. 17 (2005) 491–502.
[36] L.C. Molina, L. Belanche, A. Nebot, Feature selection algorithms: a survey and experimental evaluation, in: Proceedings of IEEE International Conference on Data Mining, IEEE Computer Society, Maebashi City, Japan, 2002, pp. 306–313.
[37] P. Somol, P. Pudil, J. Kittler, Fast branch and bound algorithms for optimal feature selection, IEEE Trans. Pattern Anal. 26 (2004) 900–912.
[38] M. Hall, Correlation-based feature selection for discrete and numeric class machine learning, in: ICML '00: Proceedings of the Seventeenth International Conference on Machine Learning, Morgan Kaufmann Publishers Inc., 2000, pp. 359–366.
[39] H.-Y. Lin, Feature selection based on cluster and variability analyses for ordinal multi-class classification problems, Knowledge-Based Syst. (2012) 11.
[40] Xin Sun, Yanheng Liu, Mantao Xu, Huiling Chen, Jiawei Han, Kunhao Wang, Feature selection using dynamic weights for classification, Knowl. Based Syst. 9 (2012).
[41] M. Petrovskiy, A game theory approach to pairwise classification with support vector machines, IEEE, Russia, 2012. p. 8.