Coordinating Discernibility and Independence Scores of Variables in a 2D Space for Efficient and Accurate Feature Selection

Juanying Xie[1(✉)], Mingzhao Wang[1], Ying Zhou[1], and Jinyan Li[2]

[1] School of Computer Science, Shaanxi Normal University, Xi'an 710062, People's Republic of China
xiejuany@snnu.edu.cn
[2] Faculty of Engineering and Information Technology, University of Technology Sydney, P.O. Box 123, Broadway, NSW 2007, Australia
Jinyan.Li@uts.edu.au

Abstract. Feature selection is to remove redundant and irrelevant features from original ones of exemplars, so that a sparse and representative feature subset can be detected for building a more efficient and accurate classifier. This paper presents a novel definition for the discernibility and independence scores of a feature, and then constructs a two dimensional (2D) space with the feature's independence as y-axis and discernibility as x-axis to rank features' importance. This new method is named FSDI (Feature Selection based on Discernibility and Independence of a feature). The discernibility score of a feature is to measure the distinguishability of the feature to detect instances from different classes. The independence score is to measure the redundancy of a feature. All features are plotted in the 2D space according to their discernibility and independence coordinates. The area of the rectangular corresponding to a feature's discernibility and independence in the 2D space is used as a criterion to rank the importance of the features. Top-k features with much higher importance than the rest ones are selected to form the sparse and representative feature subset for building an efficient and accurate classifier. Experimental results on 5 classical gene expression datasets demonstrate that our proposed FSDI algorithm can select the gene subset efficiently and has the best performance in classification. Our method provides a good solution to the bottleneck issues related to the high time complexity of the existing gene subset selection algorithms.

Keywords: Discernibility · Independence · Feature selection · Gene subset selection

1 Introduction

The fast growing of high-dimensional data sets with lots of redundant and irrelevant features brings great challenges to machine learning and data mining algorithms. Feature selection methods can choose those features which are highly correlated to labels and lowly redundant between them, without sacrificing the classification performance of the

D.-S. Huang et al. (Eds.): ICIC 2016, Part III, LNAI 9773, pp. 116–127, 2016.
DOI: 10.1007/978-3-319-42297-8_12

learning algorithm. Very often, classification models built on the selected feature subset are more accurate and easier to understand, and have a better generalization capacity, higher efficiency, reduced curse of dimensionality, and more intuitive visualization analysis [4].

A feature selection algorithm usually has two parts: feature subset search and feature subset evaluation [14]. According to the dependencies between a feature selection process and the learning algorithm, feature selection approaches can be divided into two categories: the Filters and the Wrappers [1, 11]. The filters are independent to learning algorithms, and their feature selection processes are done via an evaluation criterion which defines the feature importance without considering the learning algorithms [1]. As a consequence filters identify all the relevant features, and these features are all considered as important features to constitute the feature subset. Filters are always fast and with good generalization capability, such as Relief [10], correlation based feature selector (CFS) [6] and maximal relevance-minimal redundancy (mRMR) [15] which are the classical filter feature selection methods. Wrappers rely on the learning algorithms [11] and use the predictive accuracy of the learning algorithms on the validation datasets to test the power of the related feature subset. In general, wrappers select a sparse and representative feature subset for building a more accurate classifier. However, the computational load of wrappers is heavier than that of filters because the classification models need to be trained repeatedly in wrappers. In addition wrappers may lead to over-fitting effects on small datasets. SVM-RFE (SVM recursive feature elimination) [5] and SVM-SFS (SVM sequential forward search) [21] are typical wrappers and they have got good performance on gene expression microarray data analysis. The hybrid methods combine the filters and wrappers together to achieve a better performance. The hybrid approach has become a widely studied area for feature selection [4, 8, 21].

Gene expression data sets having tens of thousands of features but with small numbers of samples contain a high level of redundant and irrelevant gene variables for disease diagnosis purposes [3, 12, 16]. Feature selection is the primary task to analyze this type of data [4]. Time complexity bottleneck is the main issue of the available gene selection algorithms, especially for the wrappers. The cluster analysis can be applied to feature selection by choosing typical features from each cluster to construct the selected gene subset, which can partially solve the time bottleneck problem in gene selection algorithms [17, 20, 21]. However, how to detect the correct clusters is still an open question.

To select a sparse and representative feature subset and to avoid the bottleneck problems in the gene selection process, we propose a new feature selection algorithm named FSDI short for feature selection based on the discernibility and independence of a feature. FSDI defines a score for the feature discernibility and feature independence. It then uses feature independence as the y-axis and feature discernibility as x-axis to construct a two dimensional space. All features are scattered in the two dimensional space with their discernibility and independence scores as coordinates. The features at the top-right corner of the two dimensional space are those ones whose discernibility and independence are relatively large. We adopt the area of the rectangular corresponding to each feature's discernibility and independence to measure the importance of the feature. The rectangular area corresponding to the top-right corner features is

always much larger than the remaining ones, so the area for each feature can be used as a weight of the feature. Those features with larger weights are detected by FSDI to construct the sparse and representative feature subset. The classifier built on the feature subset will be more accurate. FSDI takes into account the ability of a feature to identify instances from different classes in its discernibility and the redundancy of a feature in its independence simultaneously, so that it can guarantee the selected genes are the representative ones with high relevancy to the classes and the smallest redundancy as much as possible.

On the high-dimensional gene expression data sets, we first cluster all of the genes to select those typical genes from each cluster to form the preselected gene subset. FSDI is carried out on the preselected gene subsets to get the optimal gene subsets. Experimental results on 5 classical gene expression datasets demonstrate that our FSDI method can detect gene subsets with a high efficiency, and the classifiers built on the selected gene subsets have got a better classification performance for the diagnosis purposes than those classical gene selection algorithms that are available now.

This paper is organized as follows: Sect. 2 describes the proposed FSDI method. Section 3 presents the performance results of our FSDI method on 5 high-dimensional gene expression datasets. Section 4 draws our conclusions.

2 The FSDI Algorithm

The most contribution of our FSDI is that it defines the discernibility and independence scores for each feature and constructs the two dimensional space in the feature's discernibility and independence, so that all features are scattered in the two dimensional space and the features lying at the top-right corner of the two dimensional space will be automatically detected to construct the selected feature subset to build a classifier with higher accuracy.

2.1 Preliminary Feature Selection

K-means algorithm [13] is a fast and classical clustering algorithm. It can be used to cluster big data [9, 21]. In this paper, K-means algorithm is used to cluster all genes of data into clusters of $k' = 30$. Then we use Wilcoxon Signed-rank test to measure the weight of genes and calculate the average weight of each cluster. The genes above the average weight of its cluster are preserved to constitute the preselected gene subset.

According to the principles of clustering, features in same clusters are highly similar to each other and those in different clusters are dissimilar as much as possible. Therefore, we can say that features in same clusters are highly redundant and in different clusters are relatively independent with little redundancy. Therefore the preliminary selection can reduce feature redundancy and retain feature discriminative ability, which will speed up the feature selection process and reduce the requirement of storage space and the curse of dimensionality.

Wilcoxon Signed-rank test is a nonparametric statistical method in statistic hypothesis. It can avoid the influence from parameters, so we adopt it in our research work. It is calculated in Eq. (1).

$$S(f_i) = \sum_{k=1}^{N_0} \sum_{j=1}^{N_1} \chi\left(\left(X_{j,f_i} - X_{k,f_i}\right) \leq 0\right) \tag{1}$$

where $\chi(\cdot)$ is the discriminant function, and $\chi(\cdot) = 1$ if $\left(X_{j,f_i} - X_{k,f_i}\right) \leq 0$, otherwise $\chi(\cdot) = 0$. X_{j,f_i} is the expression value of gene f_i in sample j. The number of samples in two classes of a dataset are respectively denoted by N_0 and N_1. It can be seen from Eq. (1) that the feature f_i has got good ability to detect samples from two classes when its Wilcoxon rank sum statistic value is close to 0 or $N_0 * N_1$ without considering that it has got the same value in all samples. The weight w_i of f_i is calculated by Eq. (2). The higher the value of w_i is the stronger the ability of feature f_i to discriminate samples from different classes, and the greater contribution of it to the classification task.

$$w_i = max[N_0 * N_1 - S(f_i), S(f_i)] \tag{2}$$

2.2 The Main Idea of FSDI Algorithm

Here the discernibility and independence scores of a feature are introduced, then the area of the rectangular is calculated corresponding to a feature's coordinates. FSDI algorithm selects features with much higher values of their rectangular area than the rest ones to constitute the selected feature subset.

Feature Discernibility and Independence and Importance. Suppose $\mathbf{D} = \{X_1; X_2; \cdots; X_m\} \in R^{m \times n}$ to be a train subset with m samples and n features for each. Feature discernibility and independence of feature f_i are defined as follows.

- **Feature discernibility:** we adopt Wilcoxon Signed-rank test to assess the discernibility dis_i of feature f_i and define it as the distinguishable ability of f_i to detect samples from different classes in Eq. (3). We can learn from Eq. (3) that the higher the distinguishable ability of feature f_i, is the higher value of its dis_i, that means the more importance is the feature f_i to classification.

$$dis_i = w_i \tag{3}$$

- **Feature independence:** independence of feature f_i, shown in Eq. (4), is defined as the distance of correlation between features. If a feature has the maximum discernibility value, its independence is defined by Eq. (5).

$$ind_i = \min_{j:dis_j \succ dis_i} \left(\exp\left(-r\left(f_i, f_j\right)\right)\right) \tag{4}$$

$$ind_i = \max_j \left(\exp\left(-r\left(f_i, f_j\right)\right)\right) \tag{5}$$

where r is the absolute value of Pearson correlation coefficient between features and it is calculated in Eq. (6).

$$r = \frac{\left|(\mathbf{X} - \bar{\mathbf{X}})^T(\mathbf{Y} - \bar{\mathbf{Y}})\right|}{\sqrt{\|\mathbf{X} - \bar{\mathbf{X}}\|^2 \|\mathbf{Y} - \bar{\mathbf{Y}}\|^2}} \tag{6}$$

where $\mathbf{X}, \mathbf{Y}$ indicate two feature vectors, and $\overline{\mathbf{X}}$, $\overline{\mathbf{Y}}$ are respectively the mean value of feature vector $\mathbf{X}$ and $\mathbf{Y}$.

It can be seen from (4) that the independence ind_i of feature f_i means the relevance of feature f_i with feature f_j whose discernibility is just higher than f_i. If feature f_i has got global maximum discernibility, its independence is the maximum distance of relevance with other features. The formulae (4)–(6) reveal that the feature with stronger correlation with others will obtain the weaker independence, which means the stronger the independence of a feature is, the very lower correlation of the feature with others is. This is coincidence with the real world situation.

It can be seen from (3)–(5) that the stronger distinguishability of feature f_i, is the bigger its discernibility dis_i; and the smaller the redundancy with other features, is the bigger its independence ind_i. Therefore we can construct a two dimensional co-ordinate system with features' independence as y-axis and discernibility as the x-axis, and adopt the area of the rectangle with the points (0,0) and (ind_i, dis_i) as opposite vertices to denote the importance of feature f_i. Features with large value of rectangle area possess informative information for classification whilst with little redundancy. Such kind of features are the ones we are seeking for to constitute the feature subset, which coincide with the original meaning of feature selection. The importance of feature f_i is defined in (7).

Feature Importance: the importance of feature f_i denoted as $Score_i$ is defined as the area of the rectangle surrounded by its discernibility and independence and the axes, which is calculated in Eq. (7).

$$Score_i = dis_i \times ind_i \tag{7}$$

It can be seen from formula (7) that the bigger the discernibility and independence of a feature are, the larger is the value of the importance *Score* of the feature, and at the same time the greater will the feature contribute to the classification. Calculating the importance of each feature and choosing the top features with much higher importance than the rest features to construct the feature subset, guarantee that the selected feature subset will have a better performance in classification and with very lower redundancy between features.

The FSDI Algorithm. The ideal feature subset is the one with features strong correlated to class labels whilst less redundant between features [15]. In this paper, we propose feature discernibility to measure the distinguishability of a feature between classes and feature independence to value the redundancy of a feature, and construct the two dimensional space in the feature's discernibility and independence. All features are scattered in the two dimensional space. It can be seen from the definitions of a feature's discernibility and independence in formulae (3)–(5) that the stronger the

distinguishability of a feature possesses, then the larger of its discernibility is; whilst the smaller the feature redundancy is, then the stronger of its independence is. The features with larger discernibility and independence are always scattered at the top-right corner of the two dimensional space with higher importance. FSDI will automatically detect those features with higher importance than the rest ones to construct the feature subset on which to build the classifier with more accurate. As a consequence that FSDI to some extend solved the problem of how many number of features should be selected in the feature selection algorithms.

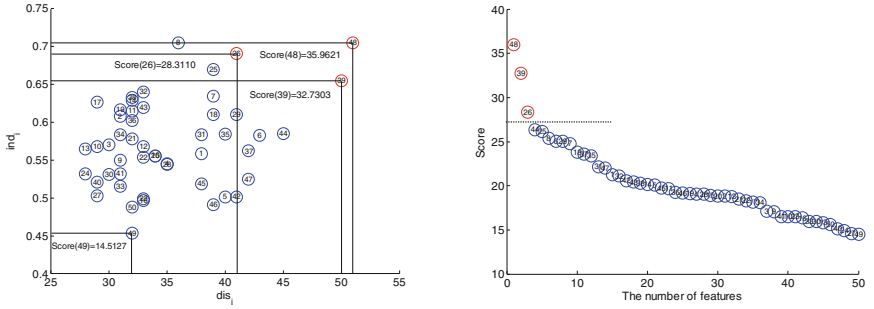
Here are the detail steps of our FSDI algorithm.
Input: train subset data ***D*** with m samples and n features for each sample, vector ***Y*** for class labels, parameter k be the number of features in the selected feature subset.
Output: the selected feature subset S.
Initialize $S = \emptyset$, F be the feature set with all features;
for i = 1 to n do
  calculate dis_i and ind_i respectively by(3) and (4) or (5);
  calculate the importance $Score_i$ of feature f_i by (7);
end for;
features are sorted in descending order in their $Score$;
add top k features with much larger $Score$ than the rest ones to S.

2.3 An Illustrating Example

Here we will test our FSDI in a random synthetic case. The synthetic dataset contains 20 samples from 2 classes, and with 50 features for each sample. We partition the synthetic dataset into train and test subset in bootstrap [7], and run our FSDI on the train subset, then build the SVM classifier on the selected feature subset. We calculate the classification accuracy in formula (8) to balance the overfitting and generalization, where M is the classifier built on the selected gene subset. We use the SVM library [2] to conduct experiments with parameter C for the linear kernel be 20. Here we do not conduct preliminary selection to features for the number of them is only 50. Figure 1(a) displays the 50 features in the two dimensional space in their discernibility and independence with their number in original dataset. Figure 1(b) shows the importance of 50 features in descending order where the y-axis is the feature importance and x-axis is the number of features.

$$Acc = 0.632 \times Acc(M)_{test} + 0.368 \times Acc(M)_{train} \tag{8}$$

It can be seen from Fig. 1(a) that the discernibility and independence of the 49th feature are very small whist the features of 48th, 39th and 26th have got the much higher discernibility and independence than the rest features. The results displayed in Fig. 1(b) disclose that the 48th, 39th and 26th features have got the highest, the second



(a) (b)

Fig. 1. The descriptions of feature importance, (a) features are scattered in their (dis_i, ind_i) in two dimensional space, (b) feature are scattered in their importance in descending order.

and the third rank of *Scores* whilst the 49th feature the smallest one. These results shown in Fig. 1(a) and (b) agree with each other.

We build the SVM classifier on the feature subset of features 48, 39 and 26. The mean accuracy of the SVM classifier is 100 %. We further analyze the performance of feature subset with top 1 and 2 features, then we respectively get the mean accuracy of 74.72 % and 89.81 %. From the analysis to the case study we can say that the definition of feature importance in this paper is reasonable and it can be used to detect those informative features for classification.

3 Experiments and Analysis

Experiments are conducted on 5 intensively studied gene expression datasets: the Colon, CNS, Leukemia, Carcinoma, and Breast Cancer datasets[1]. Table 1 describes detailed information of these datasets, where Ng and Ns denote the number of features and instances respectively. We first do the preliminary gene selection to the original genes, then run FSDI to detect the optimal gene subset from the preselected gene subset, after that we construct the classifier on the optimal gene subset.

We adopted KNN as a classification tool with K equal to 5, and compared the experimental results of our FSDI with that of the classical methods such as Weight [21], mRMR [15], ARCO [18], SVM-RFE [5] and Relief [10] when the same number of genes are selected (Table 2). Figures 2, 3, 4, 5 and 6 displayed the genes in the two dimensional space, and tagged the genes detected by our FSDI on the 5 gene expression datasets.

From the experimental results of our FSDI on these gene expression datasets shown in Figs. 2, 3, 4, 5 and 6, we can say that the features with high discernibility and independence are scattered at the top-right corner of the two dimensional space with features' discernibility and independence as coordinates. These features have got much

[1] http://levis.tongji.edu.cn/gzli/data/mirror-kentridge.html.

Table 1. The description of datasets

Gene datasets	Source	Ng	Ns
Colon	Alon *et al.*	2000	62(40 + 22)
CNS	Notterman *et al.*	7129	90(60 + 30)
Leukemia	Golub *et al.*	7129	72(47 + 25)
Carcinoma	Notterman *et al.*	7458	36(18 + 18)
Breast Cancer	Van't Veer L.J. *et al.*	24481	97(51 + 46)

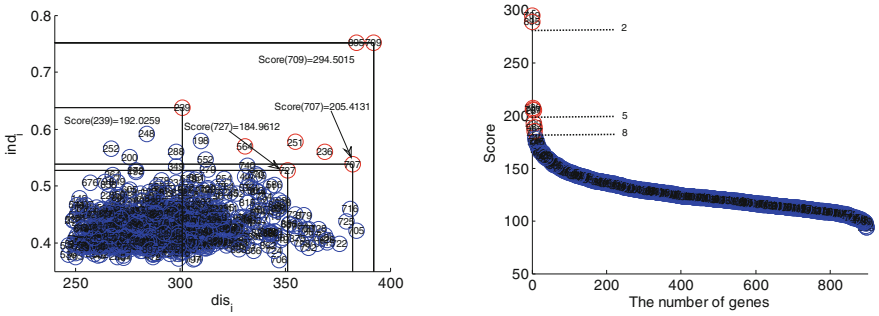


Fig. 2. The results of FSDI on Colon dataset

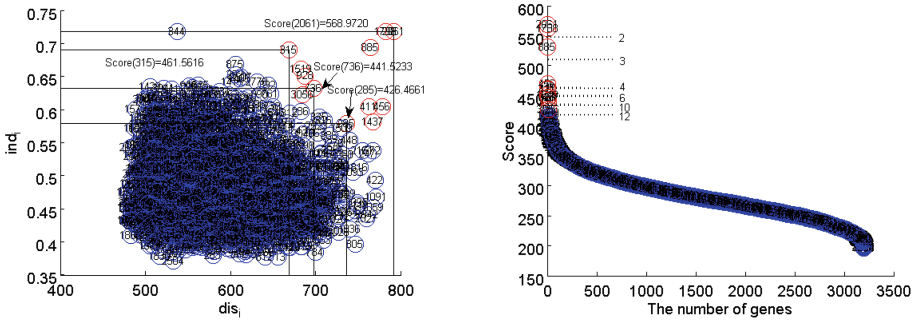


Fig. 3. The results of FSDI on CNS dataset

higher importance than the rest features which are scattered at the bottom-left corner of the two dimensional space. Our FSDI can automatically detect those features which possess much higher importance than the rest ones to construct the selected gene subsets on which to build the classifier with more accurate.

From the figures in Table 2, we can see that our FSDI dominates the other gene selection algorithms in Colon, CNS and Carcinoma datasets, and its performance on Leukemia is similar to ARCO followed by Weight, mRMR, Relief and SVM-RFE, and its performance on Breast Cancer is similar to mRMR, followed by ARCO, Weight,

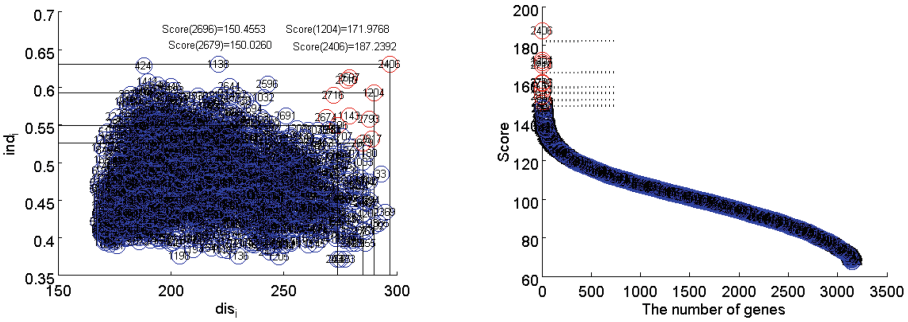


Fig. 4. The results of FSDI on Leukemia dataset

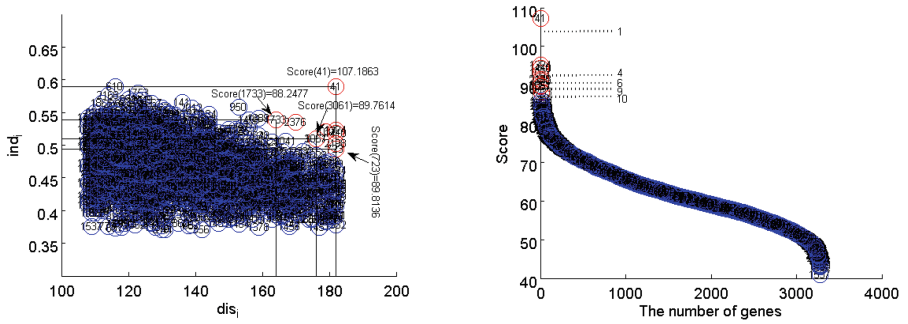


Fig. 5. The results of FSDI on Carcinoma dataset

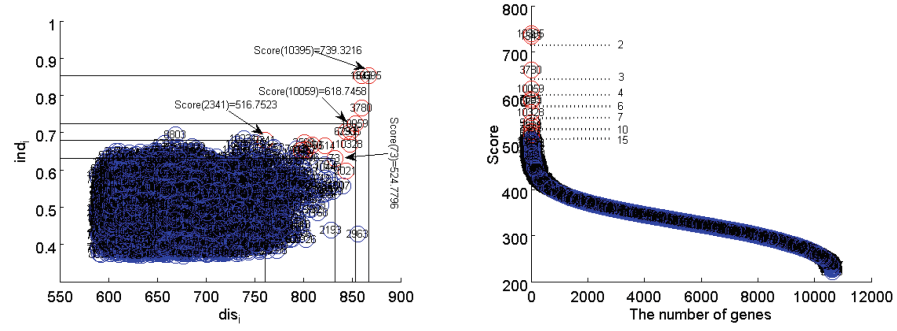


Fig. 6. The results of FSDI on Breast Cancer dataset

Relief and SVM-RFE. These results shown in Table 2 disclose that our FSDI is the best one among the compared gene selection algorithms, and also reveal that the popular gene selection algorithm SVM-RFE is the worst one among the compared gene selection algorithms. Followed our FSDI are the gene selection algorithms mRMR and

Table 2. The classification accuracy comparison of FSDI with other gene selection algorithms

Dataset	FSDI	Weight	mRMR	SVM-RFE	Relief	ARCO	Genes numbers
Colon	**0.8963**	0.5735	0.7743	0.6455	0.8010	0.8094	2
	0.8696	0.7982	0.8345	0.6606	0.8361	0.8445	5
	0.9131	0.7262	0.8361	0.8094	0.8094	0.8712	8
CNS	**0.8575**	0.7315	0.7042	0.7149	0.7423	0.8183	2
	0.9085	0.7120	0.7791	0.8230	0.7660	0.8183	3
	0.9145	0.8173	0.7957	0.8860	0.7838	0.8919	4
	0.9145	0.8173	0.8468	0.8753	0.8183	0.8860	6
	0.9323	0.8270	0.8016	0.8408	0.8242	0.9038	10
	0.9656	0.8451	0.8242	0.8242	0.8468	0.9264	12
Leukemia	**0.9346**	0.7649	0.8230	0.7374	0.8505	**0.9346**	1
	0.9256	0.7576	**0.9442**	0.8513	0.9063	0.9160	4
	0.9256	0.7762	0.9071	0.8230	0.8877	**0.9346**	6
	0.9256	0.8416	0.9071	0.7947	0.9160	0.9160	7
	0.9256	0.9249	0.9256	0.8044	0.8885	**0.9346**	8
	0.9814	**0.9814**	0.9071	0.8416	0.9071	0.9160	11
Carcinoma	**0.9864**	0.7127	0.8459	0.7369	0.9591	0.7893	1
	1	0.9847	**1**	0.9161	0.9864	0.9025	4
	1	**1**	**1**	0.9298	**1**	0.8596	6
	1	**1**	**1**	0.9298	**1**	0.9298	9
	1	0.8947	**1**	0.9298	**1**	0.8596	10
Brest Cancer	0.6922	0.6365	0.6132	0.6188	0.6245	**0.7770**	2
	0.7402	0.5902	**0.8108**	0.6161	0.6951	0.7854	3
	0.7995	0.6848	**0.8305**	0.6866	0.6640	0.7855	4
	0.7967	0.7775	**0.7967**	0.7234	0.6697	0.7544	6
	0.7601	0.7182	**0.7910**	0.6980	0.6697	0.7544	7
	0.8532	0.7200	0.8024	0.6923	0.6837	0.7996	10
	0.8237	0.7683	0.7742	0.7290	0.6753	0.8194	15

ARCO. Relief and Weight algorithms have got the similar performance when used to do gene selection for gene expression data analysis. In addition the size of the selected gene subset shown in Table 2 reveals that selected genes are sparse compared to the original ones.

From the above analysis we can state that our FSDI can find the informative genes to construct the optimal gene subset on which to build the KNN classifier with more accurate than the compared gene selection algorithms, and we can say that our FSDI can detect the gene subset with features relevant to classes and sparse and representative. Therefore we can conclude that our FSDI implement the destination of filters and wrapper simultaneously.

4 Conclusions

We have proposed a new feature selection algorithm named FSDI for informative gene selection, and defined the discernibility and independence scores of a feature to value the distinguishability and redundancy of a feature. The main contribution of this work is the construction of a two dimensional space coordinating a feature's discernibility and independence scores at the x-axis and y-axis. All the relevant features can be found and the sparse and representative feature subset can be formed by collecting those features which are always scattered at the top-right corner of the two dimensional space. Experimental results on 5 widely studied benchmark gene expression datasets demonstrate that our FSDI method has achieved better performance than the typical gene selection algorithms in terms of classification accuracy and efficiency. FSDI combines the merits of both the filter and wrapper feature selection approaches. Our method provides a good solution to the bottleneck issues related to the high time complexity of the existing gene subset selection algorithms.

Acknowledgements. We are much obliged to those who share the gene expression datasets with us. This work is supported in part by the National Natural Science Foundation of China under Grant No. 31372250, is also supported by the Key Science and Technology Program of Shaanxi Province of China under Grant No. 2013K12-03-24, and is at the same time supported by the Fundamental Research Funds for the Central Universities under Grant No. GK201503067, and by the Innovation Funds of Graduate Programs at Shaanxi Normal University under Grant No. 2015CXS028.

References

1. Blum, A.L., Langley, P.: Selection of relevant features and examples in machine learning. Artif. Intell. **97**(1), 245–271 (1997)
2. Chang, C.-C., Lin, C.-J.: LIBSVM: a library for support vector machines. ACM Trans. Intell. Syst. Technol. (TIST) **2**(3), 27 (2011)
3. Ding, C., Peng, H.: Minimum redundancy feature selection from microarray gene expression data. J. Bioinform. Comput. Biol. **3**(2), 185–205 (2005)
4. Guyon, I., Elisseeff, A.: An introduction to variable and feature selection. J. Mach. Learn. Res. **3**, 1157–1182 (2003)
5. Guyon, I., Weston, J., Barnhill, S., Vapnik, V.: Gene selection for cancer classification using support vector machines. Mach. Learn. **46**(1–3), 389–422 (2002)
6. Hall, M.A.: Correlation-based feature selection for machine learning. The University of Waikato (1999)
7. Han, J., Kamber, M., Pei, J.: Data mining: concepts and techniques: concepts and techniques. Elsevier (2011)
8. Hu, Q., Pedrycz, W., Yu, D., Lang, J.: Selecting discrete and continuous features based on neighborhood decision error minimization. IEEE Trans. Syst. Man Cybern. Part B Cybern. **40**(1), 137–150 (2010)
9. Huang, Z.: Extensions to the k-means algorithm for clustering large data sets with categorical values. Data Min. Knowl. Discov. **2**(3), 283–304 (1998)

10. Kira, K., Rendell, L.A.: The feature selection problem: Traditional methods and a new algorithm. Paper presented at the AAAI (1992)
11. Kohavi, R., John, G.H.: Wrappers for feature subset selection. Artif. Intell. **97**(1), 273–324 (1997)
12. Li, Y.-X., Li, J.-G., Ruan, X.-G.: Study of informative gene selection for tissue classification based on tumor gene expression profiles. Chin. J. Comput. Chin. Ed. **29**(2), 324 (2006)
13. MacQueen, J.: Some methods for classification and analysis of multivariate observations. Paper presented at the Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability (1967)
14. Mao, Y., Zhou, X., Xia, Z., Yi, Z., Sun, Y.: A survey for study of feature selection. Algorithm **20**(2), 211–218 (2007). (in Chinese)
15. Peng, H., Long, F., Ding, C.: Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. IEEE Trans. Pattern Anal. Mach. Intell. **27**(8), 1226–1238 (2005)
16. Shah, M., Marchand, M., Corbeil, J.: Feature selection with conjunctions of decision stumps and learning from microarray data. IEEE Trans. Pattern Anal. Mach. Intell. **34**(1), 174–186 (2012)
17. Song, Q., Ni, J., Wang, G.: A fast clustering-based feature subset selection algorithm for high-dimensional data. IEEE Trans. Knowl. Data Eng. **25**(1), 1–14 (2013)
18. Wang, R., Tang, K.: Feature Selection for Maximizing the Area Under the ROC Curve, pp. 400–405 (2009)
19. Xie, J., Gao, H.: Statistical correlation and k-means based distinguishable gene subset selection algorithms. J. Softw. **9**, 013 (2014). (in Chinese)
20. Xie, J., Gao, H.: A stable gene subset selection algorithm for cancers. In: Yin, X., Ho, K., Zeng, D., Aickelin, U., Zhou, R., Wang, H. (eds.) HIS 2015. LNCS, vol. 9085, pp. 111–122. Springer, Heidelberg (2015)
21. Xie, J., Xie, W.: Several feature selection algorithms based on the discernibility of a feature subset and support vector machines. Chin. J. Comput. Chin. Ed. **37**(8), 1704–1718 (2014). (in Chinese)