Neurocomputing 228 (2017) 270–276

Contents lists available at ScienceDirect

Neurocomputing

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

A cost-sensitive rotation forest algorithm for gene expression data classification

Huijuan Lu[a], Lei Yang[a], Ke Yan[a,*], Yu Xue[b], Zhigang Gao[c]

[a] College of Information Engineering, China Jiliang University, 258 Xueyuan Street, Hangzhou 310018, China
[b] Nanjing University of Information Science & Technology, Nanjing 210044, China
[c] College of Computer Science, Hangzhou Dianzi University, Hangzhou 310018, China

ARTICLE INFO

Keywords:
Gene expression data
Rotation forest
Cost-sensitive
Misclassification cost
Rejection cost
Test cost

ABSTRACT

Existing works show that the rotation forest algorithm has competitive performance in terms of classification accuracy for gene expression data. However, most existing works only focus on the classification accuracy and neglect the classification costs. In this study, we propose a cost-sensitive rotation forest algorithm for gene expression data classification. Three classification costs, namely misclassification cost, test cost and rejection cost, are embedded into the rotation forest algorithm. This extension of the rotation forest algorithm is named as cost-sensitive rotation forest algorithm. Experimental results show that the cost-sensitive rotation forest algorithms effectively reduce the classification cost and make the classification result more reliable.

1. Introduction

The increasingly polluted environment makes the cancer become the most common fatal disease world-widely for the current century [1]. The fast development of the Internet and database management technologies makes the automated cancer diagnosis possible [2]. In bioinformatics, various data mining and machine learning techniques are proposed to assist the cancer diagnosis in the molecular level [3–7]. With the discovery of the DNA microarray, biologists believe that the classification of gene expression data provides important information in cancer diagnosis [8–13]. The current research of the gene expression data classification problem focuses on the classification accuracy, the generalization ability, the complexity and understandability of the algorithm, the stability of the classifiers and etc. However, it is usually difficult for traditional classifiers to achieve high and stable classification results due to the three difficulties of the gene expression data classification, namely high dimension, imbalanced noisy data and small sample size [14]. Besides the classification accuracy, considering classification cost is also desired for gene expression data classification problems [15]. In this work, we take the classification cost into consideration and introduce a series of cost-sensitive learning algorithms to overcome the difficulties of the gene expression data classification.

Machine learning techniques, such as neural networks [16–18] and support vector machine [19], are widely used in gene expression data classification problems, medical diagnostic analysis, industrial data analysis and etc., because of the high classification accuracy. Decision tree (DT) is a conventional machine learning model, which generally presents a tree structure, and can be rewritten by a set of ‘if-else’ rules. Each branch of DT represents a class of sample with common characteristics in the feature space. There are many extensions of the DT algorithm, such as EG2, ID3, C4.5, CART and etc. [20].

Random forest (RF) [21] is an ensemble classifier that consists of multiple DTs through a random splitting of the feature space. The rotation forest (RoF) is developed based on RF by assembling multiple DTs [22]. It segments the feature space into subspaces, extracts the most important features from each subset and repeats the process to obtain the most distinguishable training dataset and basic classifiers for different feature subspaces. Recently, Hosseinzadeh and Eftekhari [23] proposed a high performance RoF for imbalance data classification by adding fuzzy C-mean clustering method into the RoF classification process. Fang et al. [24] utilized two improved RoF algorithms to classify highly imbalanced data. Experiments showed satisfactory results on the most widely used imbalanced measure criterion AUC.

The classification costs mainly consist of three types, namely misclassification cost, test cost and rejection cost [25]. Hybrid learning methods extend the conventional classifiers by embedding new factors into the algorithms [26,27]. By embedding the classification cost into the traditional classifiers, many cost-sensitive classification models are proposed [28]. Zidelmal et al. [29] embedded classification cost into the support vector machine (SVM) to classify the ECG beat and achieved average accuracy of 97.2% with no rejection and 98.8% for

* Corresponding author.
E-mail address: yanke@cjlu.edu.cn (K. Yan).

http://dx.doi.org/10.1016/j.neucom.2016.09.077
Received 16 May 2016; Received in revised form 25 July 2016; Accepted 5 September 2016
Available online 02 November 2016

Table 1
Average classification cost for lung dataset and ovarian dataset.

Dataset	Sample Number	30	60	90	120	150	180	210	240	270	300
lung	RoF	0.761	0.628	0.67	0.755	0.652	0.634	0.511	0.633	0.697	0.622
	C-RoF	0.661	0.75	0.629	0.551	0.587	0.548	0.505	0.482	0.477	0.452
ovarian	RoF	1.68	1.61	1.51	1.31	1.3	1.26	1.32	1.15	1.06	1.01
	C-RoF	1.13	0.84	0.82	0.78	0.69	0.62	0.57	0.568	0.57	0.562

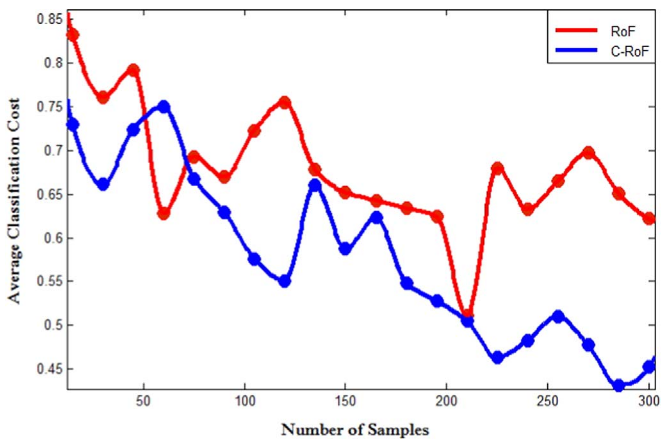

Fig. 1. Average misclassification cost in lung dataset.

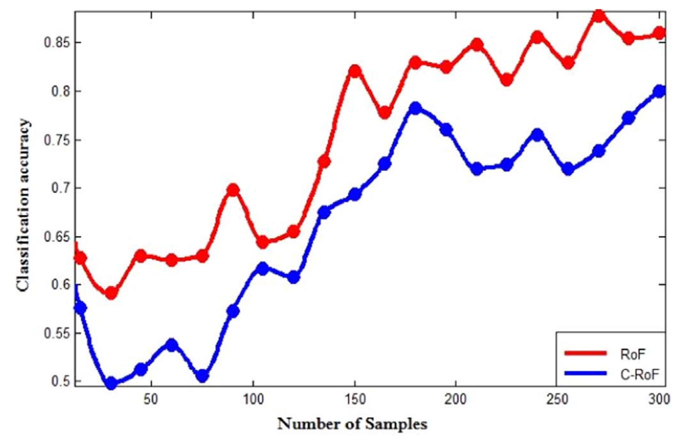

Fig. 3. Overall classification accuracy in lung dataset.

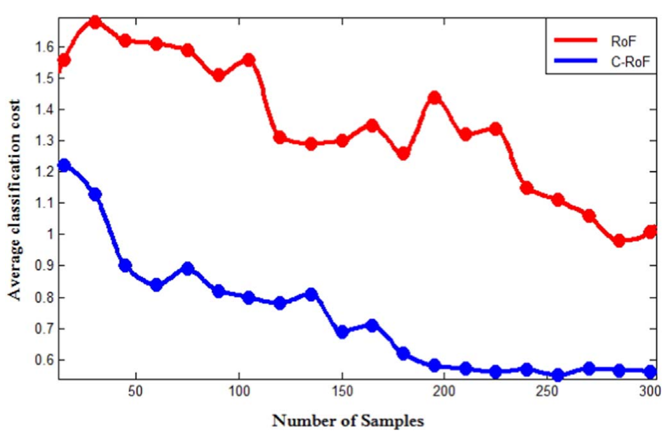

Fig. 2. Average misclassification cost in ovarian dataset.

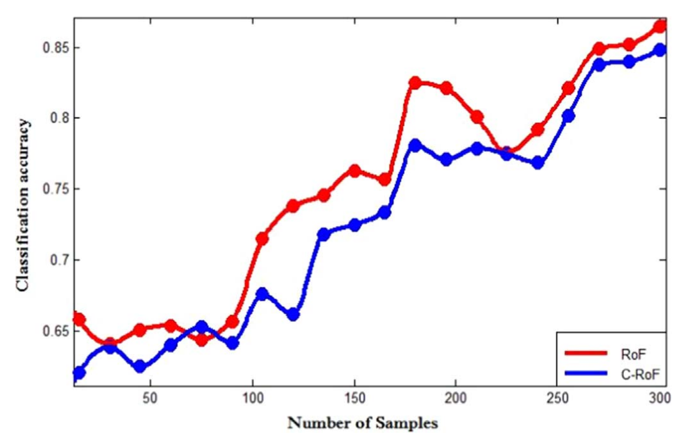

Fig. 4. Overall classification accuracy in ovarian dataset.

the minimal classification cost. Gomes et al. [30] introduced a cost-sensitive minimal learning machine for pattern classification. Lomax et al. [31] conducted a survey for cost-sensitive DT algorithms. The survey included 50 algorithms which almost covered all cost-sensitive DT algorithms that have been developed in recent years. Lu et al. [32] integrated the classification costs into the extreme learning machine (ELM) to improve the classification accuracy of the ELM. Wang et al. [33] proposed a Cost-sensitive boosting algorithm for imbalanced multi-instance datasets.

This paper proposes a cost-sensitive RoF algorithm for the classification and cost-sensitive issues of imbalanced gene expression data. First, we introduce the three main cost-sensitive factors into the gene expression data classification problem. Then, we embed the misclassification cost into the DT to improve the classification accuracy and reduce the misclassification cost of DT. Last, we embed the three factors into the RoF to form three cost-sensitive RoF algorithms. Experimental results show that the proposed three cost-sensitive RoF algorithms significantly reduce the average classification cost with guaranteed classification accuracy rates. We list the major contributions of our work as follows:

Table 2
Classification accuracy for lung dataset and ovarian dataset.

Dataset	Sample Number	30	60	90	120	150	180	210	240	270	300
lung	RoF	0.591	0.625	0.698	0.655	0.821	0.83	0.848	0.856	0.878	0.86
	C-RoF	0.498	0.538	0.573	0.608	0.693	0.782	0.72	0.755	0.738	0.8
ovarian	RoF	0.641	0.654	0.657	0.738	0.763	0.825	0.801	0.792	0.849	0.865
	C-RoF	0.639	0.64	0.642	0.662	0.725	0.781	0.779	0.769	0.838	0.848

Table 3
Average misclassification cost in lung dataset and ovarian dataset.

Dataset	Sample Number	30	60	90	120	150	180	210	240	270	300
lung	C & T-RoF	0.789	0.655	0.615	0.75	0.692	0.663	0.508	0.591	0.74	0.657
	C-RoF	0.603	0.783	0.645	0.574	0.547	0.494	0.466	0.516	0.49	0.485
ovarian	C & T-RoF	1.6	1.61	1.514	1.32	1.312	1.23	1.322	1.175	1.154	1.01
	C-RoF	1.41	1.218	1.124	1.098	1.045	0.971	0.875	0.864	0.825	0.789

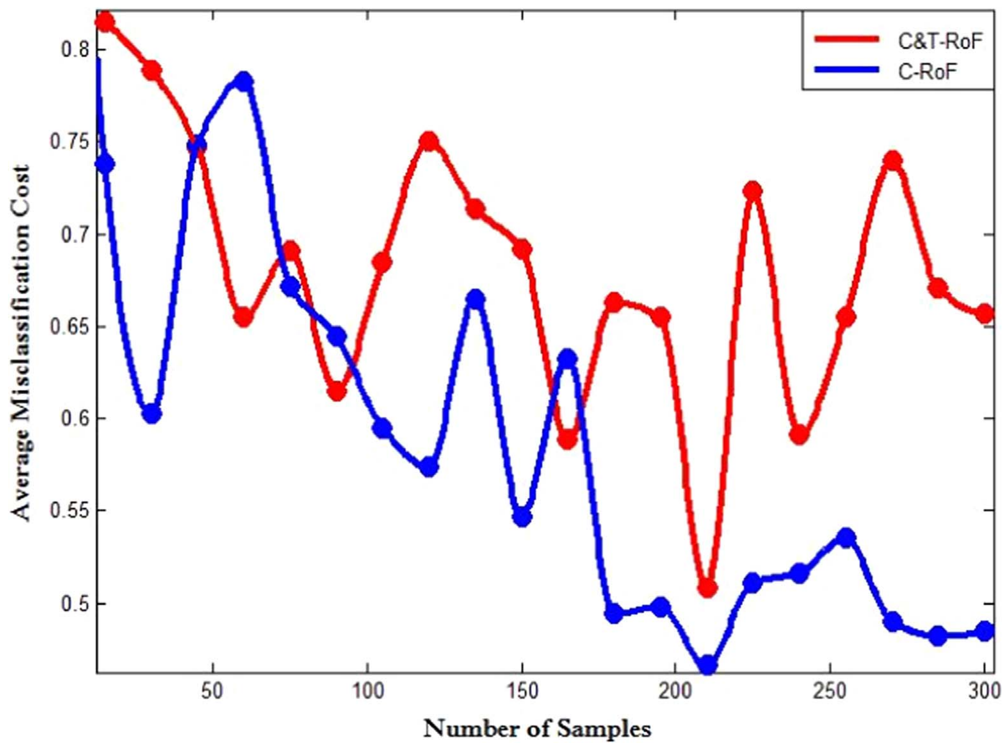

Fig. 5. Average misclassification cost in lung dataset.

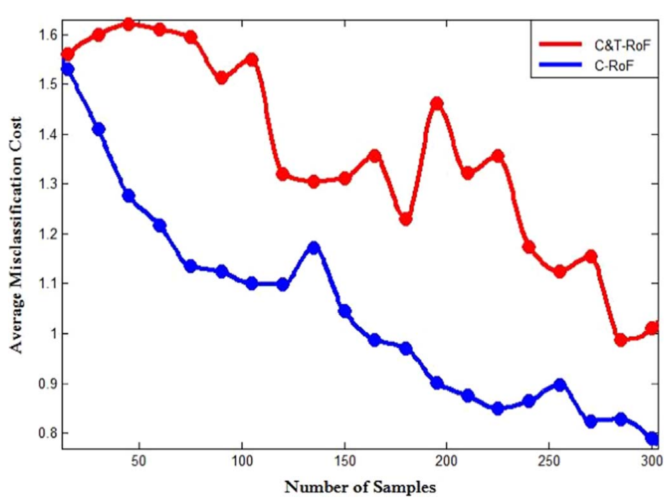

Fig. 6. Average misclassification cost in ovarian dataset.

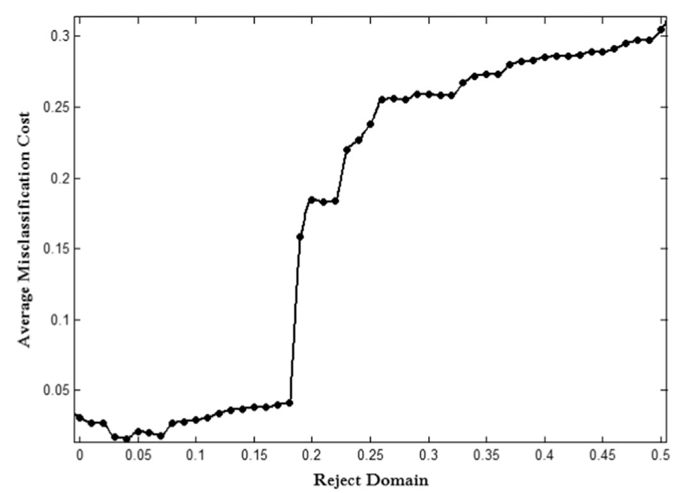

Fig. 7. The rejection domain and average misclassification cost.

- **Extending the conventional RoF algorithm by considering costs.** To address the high classification cost issue in gene expression data classification problems, we embed three different classification costs into the conventional RoF algorithm. The cost-sensitive extension of the RoF algorithm provides a new option for gene expression data classification problem, as well as broads the application scope of the RoF.
- **Minimizing the classification cost.** Experimental results show that the proposing cost-sensitive RoF has the minimal classification cost, comparing with existing approaches such as support vector machine (SVM) and extreme learning machines (ELMs).
- **Significant impacts on other related classification problems.** The successful extension of the RoF may impact the literature to encourage the development of other similar cost-

Table 4
Experimental results of different RoF algorithms.

Data	Average classification accuracy				Average misclassification cost			
	RoF	C-RoF	C & T-RoF	C & R-RoF	RoF	C-RoF	C & T-RoF	C & R-RoF
Leukemia	0.889	0.896	0.893	0.886	0.214	0.195	0.148	0.168
Breast	0.927	0.867	0.857	0.863	0.363	0.322	0.252	0.237
CNS	0.919	0.896	0.825	0.875	0.253	0.216	0.207	0.133
Heart	0.823	0.817	0.793	0.786	0.624	0.582	0.513	0.451
Colon	0.893	0.847	0.840	0.836	0.348	0.289	0.308	0.256

Table 5
Experimental results comparing C & R-RoF with existing algorithms.

Data	Overall classification accuracy				Average misclassification cost			
	C & R-RoF	ELM	SVM	CS-ELM	C & R-RoF	ELM	SVM	CS-ELM
ALL	0.865	0.876	0.853	0.811	0.207	0.447	0.857	0.376
Breast	0.863	0.746	0.784	0.728	0.237	0.756	0.908	0.652
CNS	0.875	0.887	0.903	0.837	0.133	0.554	0.734	0.256
Heart	0.786	0.743	0.795	0.625	0.451	1.054	0.947	0.683
Colon	0.836	0.895	0.872	0.825	0.256	0.401	0.618	0.279

sensitive classifiers. From the literature, we see more cost-sensitive classifiers are developing in progress and applied to various imbalance dataset problems.

2. A cost-sensitive rotation forest algorithm for gene expression data classification

The classification cost can be generally divided into three types: misclassification cost, test cost and rejection cost. In this study, we introduce all three different costs and embed them into the RoF to improve the classifier on the aspect of balancing between classification accuracy and classification cost.

2.1. Misclassification cost

The misclassification cost is the penalty value for misclassification and extended by the misclassification rate. In real-world applications, the misclassification cost determines the penalty factor in the classifiers.

Suppose that a degree-k square matrix defines a dataset with k classes of samples. Then, the cost matrix C_{ij} represents the misclassification cost of the ith class sample that is misclassified into the jth class. All diagonal values of C_{ij} are set to zeros. For example, in a binary classification problem, the misclassification cost matrix C is written as:

$$C = \begin{bmatrix} c_{00} & c_{01} \\ c_{10} & c_{11} \end{bmatrix}, \tag{1}$$

where C_{OO}(True Positive, TP) and C_{11}(True Negative, TN) represent the cost value of correct classification. C_{O1}(False Positive, FP) and C_{1O}(False Negative, FN) represent the misclassification costs. We denote $C[j]$ as the expected cost of jth class sample:

$$C[j] = \sum_{i=1}^{k} C[i, j], \tag{2}$$

The misclassification cost matrix C =[0,3;5,0] means that the cost of misclassifying the first class as the second class is 3 and the cost of misclassifying the second class as the first class is 5.

In the classification process, samples are always classified into the class of the lowest risk class by Bayesian decision theory:

$$\arg\min \sum_{j} P(j|x)\cdot C(i, j), \tag{3}$$

The class label with the lowest cost is chosen as the final category, which minimizes$R(i|x) = \sum_j P(j|x)\cdot C(i, j)$.

For multi-classification problems, the scenario is similar. Suppose x satisfies the condition $P(p|x)$ =1-$P(n|x)$. Regular classifier classifies x into class p if $P(p|x) > P(n|x)$. This result can be altered if the cost-sensitive factors are taken into consideration.

Considering the case that $P(p|x)$ =0.4, $P(n|x)$ =0.6, $C(p,n)$ =1 and $P(n,p)$ =3, the misclassification cost can be expressed as:

$$\begin{aligned} R(p|x) &= \textstyle\sum_j P(j|x)\cdot C(p, j) = p(n|x)\cdot C(p, n) = 0.6 \\ R(n|x) &= \textstyle\sum_j P(j|x)\cdot C(n, j) = p(p|x)\cdot C(n, p) = 1.2 \end{aligned} \tag{4}$$

the sample x is classified into class p, according to the principle of minimum total cost. The probability of $P(p|x)$ is 0.25, by considering the misclassification cost. Compared with the traditional classification method based on accuracy, the embedding of misclassification cost changes the classification boundary which alters the final classification results.

2.2. Test cost

Test cost refers to the expenses of obtaining attribute values. For each test (including property test, measurement test and feature test), there is a corresponding test cost. Test cost is usually used in cooperating with the misclassification cost. In modern cost-sensitive classifier design [20,25,34,35], the splitting attribute is always required to be chosen based on minimizing the sum of misclassification cost and test costs.

Embedding only misclassification cost into the classifiers may result in classification accuracy decrement because of the imbalanced splitting of the original dataset. A concrete example can be found in Section 3.1. By considering the combination of misclassification cost and test cost, the splitting attribute is selected by the minimal summation of the test cost and the misclassification cost. The tradeoffs between the misclassification cost and test cost can be balanced. For the detailed explanation of the relationship between the misclassification cost and test cost, readers may refer to the work in Ref. [25].

2.3. Rejection Cost

In real-world classification process, it is possible that the possibilities of a certain sample belonging to multiple classes are close. In this case, the misclassification risk increases without considering the rejection cost and the rejection strategies [26]. For each sample testing we set a threshold value δ to control the 'reject' operation. Similar to the misclassification cost calculation, the total cost is minimized by optimization. For an arbitrary sample t, if t satisfies:

$$P(s|t) > \max\{P(i|t)\}, \tag{5}$$

where $i \in [c_1, \cdots c_n]$ and $i \neq s$. The rejection function is defined as:

$$f(t) = P(s|t) - \max\{P(i|t)\}, \quad i \in [c_1, \cdots c_n], \quad i \neq s. \tag{6}$$

If $f(t) > \delta$, the sample is processed with regular procedures; otherwise, the sample must be processed by rejection test before the regular classification steps. We use a binary classification example to explain the steps of rejection test:

For a sample t, if $P(n|t)$ - $P(p|t) > \delta$, the label of t is n. If $P(n|t)$ - $P(p|t) < -\delta$, the label of t is p. If $|P(n|t)$ - $P(p|t)| < \delta$, the rejection strategy is executed.

The above statements can be rewritten as:

$$\bar{t}_x = \begin{cases} n, P(n|t) - P(p|t) > \delta \\ p, P(n|t) - P(p|t) < -\delta. \\ 0, |P(n|t) - P(p|t)| < \delta \end{cases} \tag{7}$$

The variable $\bar{t}_x$ represents the final class label.

In summary, for a given test sample t_y, and a combined costs matrix $C = \{C(p, n), C(n, p), C(0, n), C(0, p)\}$, where $c(p, n)$ and $c(n, p)$ represent the misclassification costs; $c(o, n)$ and $c(o, p)$ represent the rejection cost. According to the principle of risk minimization, the final class label of t_y is:

$$\bar{t}_y = \operatorname*{argmin}_i\{R(i|t)\} = \operatorname*{argmin}_i\left\{\sum_j P(j|x)\cdot C(i, j)\right\}, \tag{8}$$

where $i, j \in \{0, n, p\}$.

2.4. Embedding misclassification cost and test cost into DT

The conventional DT algorithm [34] focuses on the classification accuracy optimization, which easily leads to over-fitting for multi-class classification. In DT, each 'if-else' branch is formed by choosing a splitting attribute in the dataset which splits the dataset into two subsets. While the misclassification cost and test cost are non-negligible, we have to choose the splitting attributes that minimize the misclassification cost and embed the cost-sensitivity into DT.

The EG2 algorithm [36] employs both the information gain and testing cost factors. The information gain is represented by information cost function (ICF). Assuming that A is the splitting attribute, the information cost of A is:

$$ICF_A = \frac{2^{Gain(A)} - 1}{(C_A + 1)^{\omega}}. \quad (0 \le \omega \le 1), \tag{9}$$

where $Gain(A)$ represents the information gain. C_A represents test cost of attributes, which is determined based on experience or an expert system. Parameter ω is used to adjust the size of the information cost and the influence level of the test cost.

The EG2 algorithm is modified according to the principle of CART to include both misclassification cost and test cost. The CART decision tree usually measures the 'impurity' of nodes as the standard of choosing splitting attributes. If all data nodes come from the same class, then the 'impurity' is 0. If the classes are evenly distributed on data nodes, then the 'impurity' of nodes is big. In CART, we use the Gini coefficient [37,38] to indicate the 'impurity':

$$I(t) = 2 \times P(t) \times (1 - P(t)), \tag{10}$$

where $P(t)$ indicates the number of positive sample at node t.

The DT chooses the child node whose 'impurity' decreases the fastest to determine the split direction in t:

$$\Delta I(t) = I(t) - qI(t_L) - (1 - q)I(t_R), \tag{11}$$

where q represents the sample proportion of the left sub-node. The misclassification cost is:

$$I_C(t) = \sum_{i,j} C_{ij}p(i/t)p(j/t). \tag{12}$$

By replacing ΔI in formula (11) with I_c in formula Eq. (12), we have:

$$\Delta I_C(t) = I_C(t) - qI_C(t_L) - (1 - q)I_C(t_R). \tag{13}$$

By combining Equations from Eq. (9) to Eq. (13), we have:

$$ICF_A = \frac{2^{\Delta I_C(A)} - 1}{(C_A + 1)^{\omega}} \ (0 \le \omega \le 1). \tag{14}$$

Cost-Sensitive C4.5 (C4.5_cs) algorithm [39] introduces the weights of the samples to make the DT cost-sensitive. Weight is defined by combining the sample value and the classification cost. Samples with higher weights have higher influence values in the classification process comparing with the remaining samples.

Assume that there are N training sets T with n classes, $Cost(j)$ represents the classification cost of class j. The weights of samples can be calculated as follows:

$$weight(i) = Cost(i)\frac{N}{\sum_{j=1}^{n} Cost(j)N_j}. \tag{15}$$

C4.5_cs also uses information gains as evaluation standards. The leaf nodes are formed differently because of the sample weights. We calculate the weight of each leaf node as:

$$P_{weight}(i) = \frac{weight(i) \times N_i}{\sum_{j=1}^{n} weight(j)N_j}. \tag{16}$$

In summary, the C4.5_cs algorithm improves the classification accuracy of the classes with higher misclassification costs.

2.5. Cost-sensitive RoFs

RoFs are ensemble classifiers based on DTs. After embedding classification costs into the DT, we form the cost-sensitive RoF using the cost-sensitive DTs as base classifiers. Since there are different types of cost-sensitive DTs, we propose three approaches for cost-sensitive RoF construction:

- A misclassification cost embedded RoF (C-RoF) classifier is constructed based on the C4.5_cs.
- A misclassification and test costs embedded RoF (C & T-RoF) classifier is constructed based on the EG2.
- A misclassification and rejection costs embedded RoF (C & R-RoF) classifier is constructed based on C-RoF.

The purpose of listing three approaches is to show a progressive cost-reduction process of the RoF algorithm and compare the results of embedding different cost factors. As shown in the experimental results in Section 3, the C & R-RoF achieves the best performance with minimal classification cost and competitive classification accuracy. Both C-RoF and C & T-RoF reduce the classification cost, whereas the classification accuracy drops.

The detailed steps of constructing the proposing cost-sensitive RoF algorithms are presented as follows:

(1) Divide original feature space into K disjoint subspaces.

(2) Get bootstrap dataset using the Bagging method [40,41].
(3) Take principal component analysis (PCA) for every feature subset sample and retain all Eigen-values. Calculate the rotation matrix according to the definition in conventional RoF.
(4) Transform the training samples using rotation matrix.
(5) Choose three types of cost-sensitive DTs as the base classifier and form the three types cost-sensitive RoFs.

3. Results and analyses

We test all three cost-sensitive RoF algorithms on seven different gene expression datasets, namely lung, ovarian, acute lymphocytic leukemia (ALL), mixed lineage leukemia (MLL), breast, central nervous system (CNS) and colon with different focuses. In the experiment for the C-RoF, we focus on the average misclassification cost and overall accuracy comparison between C-RoF and conventional RoF. The C-RoF algorithm significantly reduces the misclassification cost but loses the overall classification accuracy. For C & T-RoF, both misclassification cost and test cost are embedded into the RoF to balance the tradeoff between misclassification cost reduction and classification accuracy. In the last experiment, we focus on finding the most appropriate rejection threshold. The objective of finding such a threshold is to catch the same level of classification accuracy as the original RoF with reduced misclassification cost.

3.1. Experimental results for C-RoF

In this experiment, we apply C-RoF to two datasets, namely the lung dataset and the ovarian dataset, where lung dataset has 181 samples with two different classes of the division 31:150. The ovarian dataset have 253 samples with two classes, in which 192 samples are labeled as 0 and the rest 61 samples are labeled as 1. The proportion is 3:1. The test result is evaluated by calculating the average cost taken by each classifier.

The average misclassification costs are shown in Table 1 and also depicted in Figs. 1 and 2. The overall classification accuracy rates are shown in Table 2 and also depicted in Figs. 3 and 4. All tests are repeated 30 times to obtain the average misclassification cost and classification accuracy.

By comparing Fig. 1 with Fig. 2, it shows that the average misclassification cost decreases significantly by using C-RoF. However, the overall classification accuracy of C-RoF drops after comparing Fig. 3 with 4. The decrement of the classification accuracy is due to the lack of density problem from an imbalanced data point of view when the C-RoF method chooses different splitting attributes [42]. The C-RoF method prefers splitting with attributes in low-cost direction, which alters the original way of splitting the dataset.

3.2. Experimental results for C & T-RoF

In C & T-RoF, not only the misclassification cost is embedded into the RoF algorithm, but the test cost is also embedded. In this experiment, we calculate the test cost for each attribute. The splitting attribute is selected by the minimal summation of the test cost and the misclassification cost.

According to Formula (9), we set parameterω to 0, and increase it gradually by 0.05 untilω=1. Differentω values correspond to different information cost functions. The misclassification costs can be calculated by the minimal value of ω which is obtained by the Out-of-Bag method [43]. For the simplicity of computation, we use the same ω value for all base DT classifiers of the same dataset. After cross-validating, we set ω_1 =0.6, ω_2 =0.8.

The average misclassification costs are shown in Table 3 and also depicted in Figs. 5 and 6. Comparing C & T-RoF with C-RoF, the tradeoffs between the misclassification cost and classification accuracy are further balanced, since the fundamental DT classifier takes both the misclassification cost and the test cost into account during the classification process.

3.3. Experimental results for C & R-RoF

The experiment applies the C & R-RoF method on the lung dataset first, and then expands it to other datasets. According to the suggested setting in Ref. [26], we set the rejection cost to $C(0,1)=C(0,-1)=0.2$. We gradually increase the value of rejection threshold δ (from 0 to 0.5) and set the step length to be 0.01.

From Fig. 7, we can see that the average misclassification cost is minimized when the rejection threshold δ=0.07. The cost-sensitive RoF effectively reduces the misclassification costs and shows competitive classification accuracy (Table 4).

We also compare the average classification accuracy and the misclassification cost of the final product C & R-RoF with other existing classifiers. In Table 5, we compare the C & R-RoF with ELM, SVM and cost-sensitive ELM [32]. The results show that the C & R-RoF produces the minimal misclassification cost with competitive overall classification accuracy rates.

4. Conclusion

This work has embedded the misclassification and test costs into the DT, and proposed three cost-sensitive RoF algorithms for gene expression data classification. The experimental results have shown that the C & R-RoF algorithm reduces the classification cost significantly. Moreover, the three cost-sensitive RoF algorithms improve the conventional RoF algorithm for dealing with linearly inseparable data.

Although the overall classification accuracy is improved little, the classification accuracy for the minority class might be improved due to the embedding of classification costs into the RoF. The overall classification accuracy measure rises if we add different weights to true negative classification and true positive classification [44].

Our future works include the study of integrating the cost-sensitive factors into other strong classifiers. For example, the ELM [45] is currently a popular learning method because of its fast learning speed. Our recent works show that the ELM effectively classifies the gene expression data with high accuracy rates [46,47]. We will integrate cost-sensitive factors into ELM to further improve the ELM's classification stability in the near future.

Acknowledgements

This study is supported by National Natural Science Foundation of China (No. 61272315, No. 61572164, No. 61402417, No. 61602431 and No. 61403206), Zhejiang Provincial Natural Science Foundation (No. Y1110342, No. LY15F020037) and the National Security Bureau Project (zhejiang-00062014AQ). It is also supported by the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD) and Jiangsu Collaborative Innovation Center on Atmospheric Environment and Equipment Technology (CICAEET).

References

[1] J.W. Shay, S. Bacchetti, A survey of telomerase activity in human cancer, Eur. J. Cancer 33 (5) (1997) 787–791.
[2] C. Demir, B. Yener, Automated Cancer Diagnosis Based on Histopathological Images: A Systematic Survey, Rensselaer Polytechnic Institute, 2005 (Tech. Rep).
[3] K. Polat, S. Güneş, Breast cancer diagnosis using least square support vector machine, Digit. Signal Process. 17 (4) (2007) 694–701.
[4] R. Zhang, G.B. Huang, N. Sundararajan, P. Saratchandran, Multicategory classification using an extreme learning machine for microarray gene expression cancer diagnosis, IEEE/ACM Trans. Comput. Biol. Bioinforma. (TCBB) 4 (3) (2007) 485–495.
[5] Y. Wang, I.V. Tetko, M.A. Hall, E. Frank, A. Facius, K.F. Mayer, H.W. Mewes, Gene selection from microarray data for cancer classification—a machine learning approach, Comput. Biol. Chem. 29 (1) (2005) 37–46.
[6] S.P. Deng, L. Zhu, D.S. Huang, Mining the bladder cancer-associated genes by an

integrated strategy for the construction and analysis of differential co-expression networks, BMC Genom. 16 (Suppl 3) (2015) S4.
[7] S.P. Deng, L. Zhu, D.S. Huang, Predicting hub genes associated with cervical cancer through gene co-expression networks, IEEE/ACM Trans. Comput. Biol. Bioinforma. (TCBB) 13 (1) (2016) 27–35.
[8] S. Li, X. Wu, X. Hu, Gene selection using genetic algorithm and support vectors machines, Soft Comput. 12 (7) (2008) 693–698.
[9] C.H. Zheng, D.S. Huang, L. Zhang, X.Z. Kong, Tumor clustering using nonnegative matrix factorization with gene selection, Inf. Technol. Biomed. IEEE Trans. on 13 (4) (2009) 599–607.
[10] Z. Yu, H. Chen, J. You, J. Liu, H.S. Wong, G. Han, L. Li, Adaptive fuzzy consensus clustering framework for clustering analysis of cancer data, IEEE/ACM Trans. Comput. Biol. Bioinforma. 12 (4) (2015) 887–901.
[11] Z. Yu, H. Chen, J. You, H.S. Wong, J. Liu, L. Li, G. Han, Double selection based semi-supervised clustering ensemble for tumor clustering from gene expression profiles, IEEE/ACM Trans. Comput. Biol. Bioinforma. (TCBB) 11 (4) (2014) 727–740.
[12] Z. Yu, H. Chen, J. You, G. Han, L. Li, Hybrid Fuzzy Cluster Ensemble Framework for Tumor Clustering from Biomolecular Data, IEEE/ACM Trans. Comput. Biol. Bioinforma. 10 (3) (2013) 657–670.
[13] D.S. Huang, C.H. Zheng, Independent component analysis-based penalized discriminant method for tumor classification using gene expression data, Bioinformatics 22 (15) (2006) 1855–1862.
[14] Li Shaoguang, Li Dongguang, DNA microarray dongguang technology, DNA Micro. Technol. Data Anal. Cancer Res. (2008) 1–9.
[15] Alvis Brazma, Jaak Vilo, Gene expression data analysis, FEBS Lett. 480.1 (2000) 17–24.
[16] D.S. Huang, Systematic theory of neural networks for pattern recognition, Publ. House Electron. Ind. China, Beijing 28 (1996) 323–332.
[17] D.S. Huang, Radial basis probabilistic neural networks: model and application, Int. J. Pattern Recognit. Artif. Intell. 13 (07) (1999) 1083–1101.
[18] D.S. Huang, J.X. Du, A constructive hybrid structure optimization methodology for radial basis probabilistic neural networks, IEEE Trans. Neural Netw. 19 (12) (2008) 2099–2115.
[19] K. Yan, W. Shen, T. Mulumba, A. Afshari, ARX model based fault detection and diagnosis for chillers using support vector machines, Energy Build. 81 (2014) 287–295.
[20] S.Feng, A Cost-sensitive Decision Tree under the Condition of Multiple Classes, 2015.
[21] Leo Breiman, Random Forests[J], Mach. Learn. (1) (2001).
[22] J.J. Rodriguez, L.I. Kuncheva, C.J. Alonso, Rotation forest: a new classifier ensemble method, Pattern Anal. Mach. Intell., IEEE Trans. on 28 (10) (2006) 1619–1630.
[23] M. Hosseinzadeh, M. Eftekhari, March. Improving rotation forest performance for imbalanced data classification through fuzzy clustering. In Artificial Intelligence and Signal Processing (AISP), 2015 International Symposium on pp. 35–40. IEEE, 2015.
[24] X. Fang, X. Zheng, Y. Tan, H. Zhang, Highly imbalanced classification using improved rotation forests, Int. J. Wirel. Mob. Comput. 10 (1) (2016) 35–41.
[25] X.Y. Liu, Z.H. Zhou, December. The influence of class imbalance on cost-sensitive learning: An empirical study. In Data Mining, 2006. ICDM'06.in: Proceedings of the Sixth International Conference on pp. 970–974. IEEE, 2006.
[26] Z. Yu, L. Li, J. Liu, G. Han, Hybrid adaptive classifier ensemble, IEEE Trans. Cybern. 45 (2) (2015) 177–190.
[27] Z. Yu, H. Chen, J. Liu, J. You, H. Leung, G. Han, Hybrid-Nearest Neighbor Classifier, IEEE Trans. Cybern. 46 (6) (2016) 1263–1275.
[28] S. Lomax, S. Vadera, A survey of cost-sensitive decision tree induction algorithms, ACM Comput. Surv. (CSUR) 45 (2) (2013) 16.
[29] Z. Zidelmal, A. Amirou, D. Ould-Abdeslam, J. Merckle, ECG beat classification using a cost sensitive classifier, Comput. Methods Prog. Biomed. 111 (3) (2013) 570–577.
[30] J.P.P., Gomes, A.H., Souza, Jr F., Corona, A.R.R., Neto, A Cost Sensitive Minimal Learning Machine for Pattern Classification (November)Neural Information Processing 2015 Springer International Publishing, pp. 557–564.
[31] S. Lomax, S. Vadera, A survey of cost-sensitive decision tree induction algorithms, ACM Comput. Surv. (CSUR) 45 (2) (2013) 16.
[32] H. Lu, S. Wei, Z. Zhou, Y. Miao, Y. Lu, Regularised extreme learning machine with misclassification cost and rejection cost for gene expression data classification, Int. J. data Min. Bioinforma. 12 (3) (2015) 294–312.
[33] X., Wang, S., Matwin, N., Japkowicz, X., Liu, Cost-sensitive boosting algorithms for imbalanced multi-instance datasets Advances in Artificial Intelligence, Springer Berlin Heidelberg, 2013, pp. 174–186.
[34] C.X. Ling, V.S. Sheng, Q. Yang, Test strategies for cost-sensitive decision trees. Knowledge and data engineering, IEEE Trans. on 18 (8) (2006) 1055–1067.
[35] Y. Liu, H. Lu, K. Yan, H. Xia, C. An, Applying cost-sensitive extreme learning machine and dissimilarity integration to gene expression data classification, Comput. Intell. Neurosci. 2016 (2016).
[36] K.M. Ting, An instance-weighting method to induce cost-sensitive trees, Knowl. Data Eng. IEEE Trans. on 14 (3) (2002) 659–665.
[37] L. Rutkowski, M. Jaworski, L. Pietruczuk, P. Duda, The CART decision tree for mining data streams, Inf. Sci. 266 (2014) 1–15.
[38] D. Steinberg, P. Colla, CART: classification and regression trees, Top. Ten Algorithms Data Min. 9 (2009) 179.
[39] H.Y. Lo, J.C. Wang, H.M. Wang, S.D. Lin, Cost-sensitive multi-label learning for audio tag annotation and retrieval, Multimedia IEEE Trans. on 13 (3) (2011) 518–529.
[40] L. Breiman, Bagging predictors, Mach. Learn. 24 (2) (1996) 123–140.
[41] A.A. Mikshowsky, D. Gianola, K.A. Weigel, Improving reliability of genomic predictions for Jersey sires using bootstrap aggregation sampling, J. Dairy Sci. 99 (5) (2016) 3632–3645.
[42] V. López, S. del Río, J.M. Benítez, F. Herrera, Cost-sensitive linguistic fuzzy rule based classification systems under the MapReduce framework for imbalanced big data, Fuzzy Sets Syst. 258 (2015) 5–38.
[43] H.B. Dadye, R. Rimiru, Effects of Different Pre-processing Strategies: a Comparative Study on Decision Tree Algorithms, Int. J. Digit. Content Technol. Appl. 7 (7) (2013) 939.
[44] P. Huang, C. Schechter, V. Chen, S. Hill, D. Shen, Y. Zhou, L.K. Saul, Experience in building a comparative performance analysis engine for a commercial system, Tech. Rep. CS2 (2015) 015–1014 (UCSD).
[45] G.B. Huang, H. Zhou, X. Ding, R. Zhang, Extreme learning machine for regression and multiclass classification, Syst. Man Cybern. Part B: Cybern. IEEE Trans. 42 (2) (2012) 513–529.
[46] H.J. Lu, C.L. An, E.H. Zheng, Y. Lu, Dissimilarity based ensemble of extreme learning machine for gene expression data classification, Neurocomputing 128 (2014) 22–30.
[47] H.J. Lu, E.H. Zheng, Y. Lu, X.P. Ma, J.Y. Liu, ELM-based gene expression classification with misclassification cost, Neural Comput. Appl. 25 (3–4) (2014) 525–531.

Huijuan Lu received her Ph.D. and B.S. from China University of Mining & Technology, the M.S. from Zhejiang University. Now she is the Professor of China Jiliang University. She is the executive director of CCF and the member of China cloud computing Expert Committee. She is principally engaged in cloud computing, pattern recognition, bioinformatics, data mining.

Yang lei is currently a graduate student in College of information Engineering, China Jiliang University. He received his B.S. from Hangzhou Dianzi University in 2015. His research interests are in the fields of pattern recognition, machine learning and data mining.

Dr. Ke Yan completed both the Bachelor's and Ph.D. degree in National University of Singapore (NUS). He received his Ph.D. certificate in computer science in 2012 under the supervision of Dr. Ho-Lun Cheng. During the years between 2013 and 2014, he was a post-doctoral researcher in Masdar Institute of Science and Technology in Abu Dhabi, UAE. Currently, he serves as a lecturer in China Jiliang University, Hangzhou, China. His main research field includes computer graphics, computational geometry, data mining and machine learning.

Yu Xue was born in 1981. He is a member of IEEE (92058890), ACM (2270255), and CCF (E200029023M). He received the Ph.D. degree from College of Computer Science and Technology, Nanjing University of Aeronautics & Astronautics, China, in 2013. He is a lecturer in the School of Computer and Software, Nanjing University of Information Science and Technology. He has published nearly twenty journal and conference papers. His research interests include computational intelligence, internet of things and electronic countermeasure.

Zhigang Gao received the Ph.D. degree from the College of Computer Science, Zhejiang University, Hangzhou, China in 2008. He is a teacher in the College of Computer Science, Hangzhou Dianzi University, Hangzhou, China. His current research interests are pervasive computing, Cyber-Physical Systems, and automotive electronic systems.