Measure Oriented Cost-Sensitive SVM for 3D nodule detection

Peng Cao, Dazhe Zhao, and Osmar Zaiane

Abstract—The class imbalance issue occurs when training a computer-aided detection (CAD) system for nodules. This imbalance causes poor prediction performance for true nodules. Moreover, the misclassification costs are different between two classes and high sensitivity of true nodules is essential in the detection. In order to eliminate or reduce the false positives while keeping high sensitivity, we present an effective wrapper framework incorporating the evaluation measure of imbalanced data into the objective function of cost sensitive SVM. We improve the performance of classification by simultaneously optimizing the best pair of misclassification cost parameter, feature subset and intrinsic parameters. We evaluated the method on a 3D Lung nodule dataset, showing that the proposed method outperforms many other exiting common methods, as well as specific imbalanced data learning methods, which indicates the effectiveness of our method on the imbalanced and unequal misclassification cost data classification.

I. INTRODUCTION

Lung cancer is one of the main public health issue in developed countries [1], and early detection of pulmonary nodules is an important clinical indication for early-stage lung cancer diagnosis. Computer aided detection (CAD) can provide initial nodule detection which may help expert radiologists in their decision making. A CAD scheme for nodule detection in CT (Computed Tomography) can be broadly divided into a nodule identification step and a false-positive reduction step. For finding the suspicious nodules, the initial detection of the CAD requires high sensitivity, and so, it produces a number of false positives. Since the radiologists must examine each identified object, it is highly desirable to reduce these false positives (FPs) while retaining the true positives (TPs). Therefore significant efforts are needed in order to further improve the performance of current CAD schemes for nodule detection [2].

The false-positive reduction (FPR) step is a critical part in the Lung nodule detection system [3, 4]. It is a classification task. The most significant problems in the FPR is that the two classes are skewed and have unequal misclassification costs in the training set. Typically the minority class carries a higher cost of misclassification, making common classification methods inappropriate. This is a typical "class imbalance problem" [5]. Class imbalanced data has detrimental effects on the performance of conventional classifiers [5]. However, in the potential nodule classification, the problem has attracted less attention. Only few works consider this problem using simple solutions [3, 6-8]. There is still room for improvement on the FPR.

The results obtained in [7] show that cost-sensitive SVM (CS-SVM) trained with imbalanced data sets achieves promising results in terms of sensitivity and specificity, by means of adjusting the misclassification cost of false positives versus false negatives. However, there are some challenges with respect to the training of CS-SVM. The misclassification costs play a crucial role in the construction of the classification model. However, the appropriate misclassification costs cannot be apriori set. It is not correct to simply set the ratio cost to the reverse of the sizes of two classes, as it is often done. Beside the cost, there are two other problems confronting SVM: how to choose the optimal feature subset for SVM, and how to set the best intrinsic parameters including the regularization parameter C and the kernel function parameter. The issue of high-dimensionality poses additional challenges when dealing with class-imbalanced prediction, resulting in influencing the performance of the CS-SVM. Thus, it is important to select features that lead to a higher separability among the unequal classes [9]. In addition, the intrinsic parameters of SVM also influence the classification performance. Furthermore, the feature subset choice influences the appropriate intrinsic parameters as well as the misclassification cost and vice versa. Therefore, these factors need to be obtained simultaneously. The significance of the scheme has two questions to fix: how to optimize these factors simultaneously and what evaluation criteria to use for guiding their optimization. These two issues are our key step for improving the CS-SVM in the context of the suspicious nodule classification. Our main contributions in this paper are centered around the questions above.

In order to solve the challenges above, we design a novel CS-SVM driven by performance measures. Particle swarm optimization (PSO) is employed as the optimization strategy due to its fast and effective solution space exploration [10]. The measure oriented scheme optimizes the factors (cost, feature subset and intrinsic parameters) simultaneously. Furthermore, we use different performance evaluation metrics as objective functions like the G-mean [11], AUC (Area Under ROC Curve) [12]. Different metrics can reflect different aspect performance of SVM classifier.

II. MEASURE ORIENTED COST SENSITIVE LEARNING

A. Cost-Sensitive SVM

The cost-sensitive version of SVM (CS-SVM) [13] is normally a good solution to address the above problem. CS-SVM is formulated as follows:

$$Min \ \frac{1}{2}\|\boldsymbol{w}\|^2 + C_+\sum_{i:y_i=+1}\xi_i + C_-\sum_{j:y_j=-1}\xi_j$$
$$s.t. \ \ y_i[(\boldsymbol{w}^T\boldsymbol{x_i})+b]\geq 1-\xi_i \ \ i=1,\cdots \tag{1}$$
$$\xi_i \geq 0$$

where the C_+ is the higher misclassification cost of the positive class, which is the primary interest, while C_- is the

lower misclassification cost of the negative class. Using the different error costs for the positive and negative classes, the hyperplane could be pushed away from the positive instances. In this article, we fix $C_- = C$ and $C_+ = C \times C_{rf}$, where C and C_{rf} are respectively the regularization parameter and the ratio misclassification cost factor. We choose the Radial Basis Function (RBF) as kernel function of SVM.

B. Measure optimized Cost-Sensitive SVM

This algorithm uses a wrapper approach to perform the search for the potentially optimal misclassification cost, intrinsic parameters, and feature subset of CS-SVM simultaneously in the parameter space under the guidance of performance evaluation measure.

In this paper, for the multivariable optimization, especially the hybrid multivariable, the best methods are swarm intelligence technique. We choose the particle swarm optimization (PSO) as our optimization method because it is very mature and easy to implement. In addition, many experiments claim that PSO has equal effectiveness but superior efficiency over the GA [14]. PSO optimizes an objective function by a population-based search. The population consists of potential solutions, named particles. These particles are randomly initialized and move across the multi-dimensional search space to find the best position according to an optimization function. During optimization, each particle adjusts its trajectory through the problem space based on the information about its previous best performance (personal best, *pbest*) and the best previous performance of its neighbors (global best, *gbest*). Eventually, all particles will gather around the point with the highest objective value. The position of individual particles is updated as follows:

$$x_i^{t+1} = x_i^t + v_i^{t+1} \qquad (2)$$

with v, the velocity calculated as follows:

$$v_{id}^{t+1} = w \times v_{id}^t + c_1 \times r_1 \times (pbest_{id}^t - x_{id}^t) + c_2 \times r_2 \times (gbest^t - x_{id}^t) \qquad (3)$$

where v_i^t indicates velocity of particle i at iteration t, w indicates the inertia factor, C_1 and C_2 are the cognition and social learning rates, which determine the relative influence of the social and cognition components. r_1 and r_2 are uniformly distributed random numbers between 0 and 1, x_i^t is the current position of particle i at iteration t, $pbest_i^t$ indicates best of particle i at iteration t, $gbest^t$ indicates the best of the group.

(1) Particle design

The misclassification cost, intrinsic parameters and feature subset for SVM need to be searched at the same time, thus the solution in PSO includes three parts: the ratio misclassification cost, the intrinsic parameters, and the feature subsets. For feature subset, each feature is represented by a 1 or 0 for whether it is selected or not. **Fig. 1** illustrates the mixed solution representation in the PSO.

C_{rf}	C	γ	f_1	...	f_n

Fig. 1 Solution representation

The variables needed to be optimized are enormous and mixed, since the costs and parameters we intend to optimize are continuous while the feature selection is discrete. PSO

was originally developed for continuous valued spaces, the discrete PSO [15] can solve the discrete variables. The major difference between the discrete PSO and the original version is that the velocities of the particles are rather defined in terms of probabilities that a bit will change to one. Using this definition a velocity must be restricted within the range [0, 1], to which all continuous values of velocity are mapped by a sigmoid function:

$$v''_i = sig(v_i^t) = \frac{1}{1 + e^{-v_i^t}} \qquad (4)$$

Equation 4 is used to update the velocity vector of the particle while the new position of the particle is obtained using **Equation 5**.

$$x_i^{t+1} = \begin{cases} 1 & if \quad r_i < v''_i \\ 0 & otherwise \end{cases} \qquad (5)$$

Where r_i is a uniform random number in the range $[0,1]$.

(2) Fitness function

Evaluation measures play a crucial role in both assessing the classification performance and guiding the classifier modeling. For imbalanced datasets, the average accuracy is not an appropriate evaluation metric. We used the G-mean and AUC as the fitness function to optimize the CS-SVM separately. These two different evaluations reflect different aspect of the classifier. The AUC concerns the ranking ability and the G-mean concerns the two accuracies of both classes at the same time ($G - mean = \sqrt{Sensitivity * Specificity}$).

(3) The procedure for the proposed MOCS-SVM

The detailed algorithm about the measure oriented cost sensitive SVM, MOCS-SVM, is shown in **Algorithm 1**.

Algorithm 1 MOCS-SVM

Input: Training set *D*; termination condition *T*; population siz *SN*; metric *E*; *NumFolds*(3)
Randomly initialize particle population positions and velocities (including cost, intrinsic parameters, and feature subset)
repeat
  foreach particle *i*
    Construct the D_i with the feature selected by the particle *i* (x_i)
    for *k*=1 to *NumFolds*
      Separate D_i randomly into Trt_i^k (80%) for training *and* Trv_i^k (20%) for validation
      Train CS-SVM with ratio cost and intrinsic parameters optimized by the particle *i* on the Trt_i^k
      Evaluate the cost sensitive classifier on the Trv_i^k and obtain the value M_i^k based on evaluation metric *E*
    end for
    M_i=average(M_i^k), and assign the fitness of particle *i* (x_i) with M_i
    if *fitness* ($pbest_i$) <= *fitness* (x_i)   **then** $pbest_i = x_i$
  end foreach
  set *gbest* as best *pbest*
  foreach particle *i*
    update *velocity_i* and *position_i* with Eq. 2 - 5.
end foreach
until *T*
output optimal parameters, ratio cost and feature subset of *gbest*

It is a wrapper framework for empirically discovering the potential misclassification cost ratio, intrinsic parameters and feature subset for CS-SVM oriented by the measure (G-mean and AUC). The 3-fold cross validation is applied to evaluate classification performance for each potential solution of particles to avoid any estimation biases. The

averaged performance measure is calculated as the fitness value of each solution in the particle.

III. EXPERIMENTAL STUDY

A. Potential Nodule Candidates Detection

Our database consists of 98 thin section CT scans with 106 solid nodules, obtained from Guangzhou hospital. These databases included different sizes nodules, annotated by three expert radiologists. The nodule locations of these scans are marked by expert radiologists. For obtaining the candidate nodules, we employ the 3D hessian filter to detect the candidate nodule VOI (Volume of Interest) and use the 3D region growing to get the core region [16]. **Fig. 2** shows some example result images of candidate VOI detection. We obtained 95 true nodules as positive class and 592 non-nodules as negative class from the total CT scans.
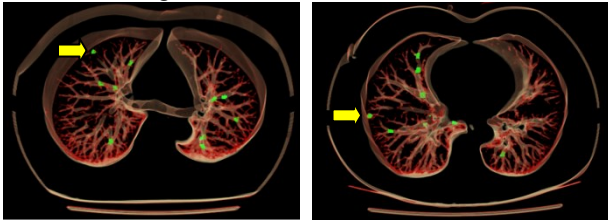

Fig. 2 Potential nodule initial detection. TP indicated by arrow, other spots are FP

B. Feature extraction

In order to more accurately identify true or false positive nodules, we calculated multiple types of image features for each nodule candidate: intensity, shape and gradient. This feature extraction process generated 43 image features. This section gives a brief introduction to the features we have extracted. All the features are scaled into [-1,1] in the dataset.

(1) Intensity feature

The nodules often have higher gray values than parts of vessels misidentified as nodules; and the intensity distribution of nodule's CT appearance can be approximated by a Gaussian function. Therefore, we use the statistical global gray feature and the radial distribution feature to describe the gray distribution within the candidate objects. The gray value within the objects was characterized by use of seven statistics (mean, variance, max, min, skew, kurtosis, entropy).

In order to capture the volume intensity distribution along the radial direction of the nodule candidates, we extract the feature of radial intensity distribution (*RIS*). The vector of *RIS* feature for nodule object should be decreasing, while for a non-nodule it changes irregularly. It is computed as follows:

$$RIS(r_{i+1}, r_i) = \sum_i I(x_i, y_i, z_i) \tag{6}$$

Where $I(x_i, y_i, z_i)$ is the gray value in the position of (x_i, y_i, z_i), and $(x_i, y_i, z_i) \in \{(x_i, y_i, z_i) \mid r_i < \sqrt{(x_i - x_c)^2 + (y_i - y_c)^2 + (z_i - z_c)^2} < r_{i+1}\}$, (x_c, y_c, z_c) is the coordinate of the center.

We divide the volume into five sub-volumes with equal radius. The average intensity values of $RIS(r_{i+1}, r_i)$ $(i=0,\ldots,4)$ are calculated in each sub-volume and normalized.

(2) Shape feature

Based on the fact that an isolated nodule or a nodule attached to a blood vessel is generally either depicted as a sphere or has some spherical elements, while a blood vessel is usually oblong, we attempt to distinguish true nodules from false ones by calculating the volumetric shape index (SI), curvedness (CV) [16], and other explicit shape features to characterize the 3D shape. SI represents the local shape feature at each voxel while being less sensitive to the image intensity. It can be defined as:

$$SI(p) = \frac{1}{2} - \frac{1}{\pi} \arctan \frac{k_1(p) + k_2(p)}{k_1(p) - k_2(p)} \tag{7}$$

where $k_1(p)$ and $k_2(p)$ are principal curvatures at voxel p. The curvedness quantifies how highly curved a surface is. It can be calculated as follows:

$$CV = \sqrt{\frac{k_1^2 + k_2^2}{2}} \tag{8}$$

The SI and CV features are calculated based on each pixel, and they are further characterized by seven statistical operations introduced above. Some other explicit shape features were also extracted including the volume, the surface area and compactness.

(3) Gradient distribution feature

The true nodules have a high concentration because they grow from the center to surround, thus nodules have high concentration of gradient vector. We utilize the gradient concentration (GC) feature to characterize the degree of convergence of the gradient vectors at voxel p [17]. It is defined by:

$$GC(p) = \frac{1}{D} \sum_{i=1}^{D} e_i^{\max}(p), \quad e_i^{\max}(p) = \max_{R_{\min} \leq n \leq R_{\max}} \left\{ \frac{1}{n+1} \sum_{l=0}^{N} \cos\theta_{il}(p) \right\} \tag{9}$$

where D is the number of the symmetrically direction vectors d^l originating from p. The angle $\cos\theta_{il}(p)$ is calculated between d^l and g_i^l, where g_i^l is the gradient vector located at distance l from p in direction d^l. We also calculate the gradient field strength at each pixel. The concentration and strength features based on every pixel are further characterized by seven statistical operations.

C. Potential nodules candidates classification

In this experiment, for the PSO setting in MOCS-SVM, the initial parameter values of it in our proposed method were set according to the conclusion drawn in [18]. The parameters used were: C_1=2.8, C_2=1.3, w=0.5. The particle number was set by 30, and the termination condition was set by 500 cycles. Along with these parameters in PSO, the other parameters are the upper and lower of limit parameters to be optimized. The ranges for C and γ are based on the one as recommended in [19]. The range of C and γ are $(2^{-5}, 2^{15})$ and $(2^{-15}, 2^3)$, respectively. The range of ratio misclassification cost factor C_{rf} was empirically set $[1, 100 \times N_{neg} / N_{pos}]$. All the experiments are conducted by 10-fold cross validation. LibSVM is used as the implementation of SVM [19].

(1) Experiment I

In this experiment, the comparison is conduct between the intermediate or basic methods, such as basic SVM, basic CS-SVM, optimized CS-SVM with grid search, and our method MOCS-SVM with/without the feature selection guided by G-mean (MOCS-SVM$_{GM}$) and AUC (MOCS-SVM$_{AUC}$). All SVM models in this experiment use

the RBF kernel. For basic SVM and CS-SVM, the intrinsic parameters are chosen with default values (C=1 and γ =1), ratio cost $C_{rf}=N_{neg} / N_{pos}$. For the optimized CS-SVM with grid search, we also need to treat this misclassification cost ratio as a parameter to optimize. Since it is not feasible to use a triple circulation for optimizing the best parameters, we optimize the best parameter pair(C and γ) firstly, then optimize the cost ratio locally to maximize G-mean based on the best parameter pair(C and γ) obtained before.

Table 1. Experiment results with respect to AUC, G-mean (GM) and the feature size (Fea)

	SVM	CS-SVM	Grid CS-SVM	MOCS-SVM$_{GM}$		MOCS-SVM$_{AUC}$	
					FS		*FS*
AUC	0.751	0.816	0.854	0.879	0.929	0.915	**0.937**
GM	0.526	0.648	0.739	0.834	**0.878**	0.827	0.855
Fea	43			43	22	43	24

From the result in **Table 1**, we found that simultaneously optimizing the feature subset, parameter and cost ratio generally helps CS-SVM, regardless of feature selection or not. Meanwhile, under the condition where the feature selection is not carried out, we found the optimization for all the factors simultaneously with PSO outperforms the optimization with grid search, which optimizes the intrinsic parameter firstly, then search the optimal misclassification cost parameter based on the best intrinsic parameters. It lacks many potential parameters pairs not to be searched in the parameter space. Hence, it shows that the parameters need to be searched at the same time. As expected using different measure for guiding optimization results in different feature subset. Because different measures have different biases, they select different feature subset to obtain the best performance.

(2) Experiment II

We conduct the comparison between our method and the other state-of-the-art approaches for dealing with the imbalanced data, such as AdaCost [20], random under-sampling (RUS) [5], SMOTE (SM) [21] and SMOTEBoost (SMB) [22]. All the three sampling methods re-balance the two classes until the sizes of both classes are the same. The ratio cost of AdaCost is set to the ratio between the sizes of two classes. The results are shown in **Table 2**. We can see that MOCS-SVM produced the best results among all the tested methods for imbalanced data learning. It means that our method can be applied for FPR on the nodule detection.

Table 2. Experiment result with respect to the value of AUC and G-mean

	AdaCost	RUS	SM	SMB	MOCS-SVM$_{GM}$	MOCS-SVM$_{AUC}$
AUC	0.785	0.703	0.848	0.855	0.929	**0.937**
GM	0.662	0.590	0.726	0.748	**0.878**	0.855

IV. CONCLUSION

The false positive reduction is a class imbalance task in the Lung nodule detection. In this paper, we propose a wrapper paradigm oriented by the performance measure as objective function with respect to misclassification cost, feature subset and intrinsic parameter of SVM. The optimization processing is through an effective swarm intelligence technique, PSO. Our measure optimized

framework could wrap around an existing cost-sensitive classifier. Through theoretical justifications and empirical studies, we demonstrated this method to be effective in the FPR issue. This method can be applied on other computer aided detection, such as for breast or colon. Due to the nature of PSO, searching of parameters might be time-consuming. Since this searching is usually an off-line procedure, the time for learning is not a crucial issue.

REFERENCES

[1] R.T. Greenlee, T. Murray, S. Bolden, P.A. Wingo, Cancer statistics, 2000, CA Cancer J Clin 2000;50:7–33.

[2] Q. Li, Recent progress in computer-aided diagnosis of lung nodules on thin-section CT, Computerized Medical Imaging and Graphics 2007, 31(4-5), 248-257.

[3] L. Boroczky, L.Z. Zhao. & K.P. Lee, Feature Subset Selection for Improving the Performance of False Positive Reduction In Lung Nodule CAD, IEEE Transactions On Information Technology In Biomedicine, 2006, vol. 10, NO. 3.

[4] K. Suzuki, S.G. Armato, F. Li, S. Sone & K. Doi, Massive training artificial neural network for reduction of false positives in computerized detection of lung nodules in low-dose computed tomography, Med. Phy., 2003, vol. 30, pp. 1602–1617.

[5] H. He, E.A. Garcia, Learning from imbalanced data, Knowledge and Data Engineering, IEEE Transactions on 21,2009, 1263-1284.

[6] J. Dehmeshki, J. Chen, M.V. Casique, & M. Karakoy, Classification of lung data by sampling and support vector machine, IEEE EMBS annual international conference, 2004, 3194-3197.

[7] P. Campadelli, E. Casiraghi, & G. Valentini, Support vector machines for candidate nodules classification, Neurocomputing 68, 2005, 281–289.

[8] M. Dolejsi, J. Kybic, M. Polovincak. & S. Tuma, Reducing false positive responses in lung nodule detector system by Asymmetric AdaBoost, ISBI. 2008, 656-659.

[9] J. Van Hulse, T. Khoshgoftaar, A. Napolitano, R. Wald, Feature selection with high-dimensional imbalanced data, in: ICDM, pp. 507-514, 2009.

[10] J. Kennedy, R.C. Eberhart, Particle swarm optimization, IEEE Int. Conf. Neural Networks, 1995, pp.1942–1948.

[11] M. Kubat, R. C. Holte, & S. Matwin, Machine learning for the detection of oil spills in satellite radar images, Machine learning, 1998, 30(2), 195-215.

[12] T. Fawcett, ROC graphs: Notes and practical considerations for researchers, Machine Learning 2004, 31, 1-38.

[13] K. Veropoulos, C. Campbell, N. Cristianini, Controlling the sensitivity of support vector machines, Proceedings of the International Joint Conference on AI, 1999: 55–60.

[14] D. Martens, B. Baesens, T. Fawcett. Editorial Survey: Swarm Intelligence for Data Mining, Machine Learning 82(1):1-42, 2011.

[15] M.A. Khanesar, M. Teshnehlab, M.A. Shoorehdel, A novel binary particle swarm optimization. In Control & Automation. MED 07. Mediterranean Conference on, 2007, pp. 1–6, Athens.

[16] Q. Li, F. Li & K. Doi, Computerized detection of lung nodules in thin-section CT images by use of selective enhancement filters and an automated rule-based classifier, Academic radiology, 2008 15(2): 165.

[17] H. Kobatake and M. Murakami, Adaptive filter to detect rounded convex regions: Iris filter, in: ICPR, vol.II, 1996, pp. 340–344.

[18] A. Carlisle, G. Dozier, An Off-The-Shelf PSO. Particle Swarm Optimization Workshop, 2001, pp. 1–6.

[19] Hsu CW, Chang CC, Lin CJ. A Practical Guide to Support vector Classification, National Taiwan University Technical Report, 2003.

[20] W. Fan, S.J. Stolfo, J. Zhang, & P.K. Chan, AdaCost: Misclassification Cost-Sensitive Boosting, In: ICML, pp. 97-105, 1999.

[21] N.V. Chawla, K.W. Bowyer, L.O. Hall & W.P. Kegelmeyer, SMOTE: Synthetic minority over-sampling technique. Journal of Artificial Intelligence Research, 2002, pp. 341–378.

[22] N.V. Chawla, A. Lazarevic, L.O. Hall, K.W. Bowyer, SMOTEBoost: Improving Prediction of the Minority Class in Boosting. in: PKDD, 2003, pp. 107-119.