Available online www.jocpr.com

Journal of Chemical and Pharmaceutical Research, 2013, 5(12):183-187

Research Article

ISSN : 0975-7384
CODEN(USA) : JCPRC5

Computerized distinction of benign and malignant pulmonary nodules on PET-CT imageology character

Yan Qiang, Yanbo Ma, Juan-juan Zhao* and Jin Wang

College of Computer Science and Technology, Taiyuan University of Technology, Taiyuan, China

---

ABSTRACT

In order to improve the accuracy of the solitary pulmonary nodule diagnosis with medical signs in medical imaging diagnostics, a novel computer-aided classification method is developed. The method use features from Computed Tomography (CT) images combined with Standard Uptake Value (SUV) values in Positron Emission Tomography (PET) images to build a Support Vector Machine (SVM)classifier model. Using particle swarm optimization on SVM parametric search,thus choose the most appropriate parameters. After that will get the appropriateSVM classification model. The experimental results show that usesthetwo kinds of featuresin this study can achieve a high classification accuracy. However, the SVM classification model which after parameters optimized has higher classificationaccuracy.This method can avoid the randomness of human choice and provides a theoretical basis for the main features selected when doctors diagnosis of pulmonary nodules.

Keywords: PET/CT, Solitary PulmonaryNodule, SupportVectorMachine, ParticleSwarmOptimization, Computer-Aided

---

INTRODUCTION

Solitary pulmonary nodule(SPNs) [1]which is a single, oval and maximum diameter is no more than 30 mm nodule in lung parenchyma. Early resection in patients with malignant lung nodules five-year survival rate can be higher than 60% [2], but due to the diagnosis is not clear enough, half of the removal nodules are benign. Therefore, accurate diagnosis of benign and malignant solitary pulmonary nodules is very critical.

CT is the conventional method; the sensitivity is up to 97%, but the specificity only 70%. PET has a higher accurate rate whendiagnosis of pulmonary nodules, but lower sensitivity and specificity[3,4]. Combined the two characteristics namely PET / CTimagecan results to better diagnosis effect. But extract the CT images and PET images at the same time making the amount of image data to a sharp growth and lung disease has so many types that the PET and CT imaging presents are very complex[5]. In this case, doctor's diagnosis virtue of experience has a strong subjectivity. Therefore, it is necessary to resort to the computer-aided diagnosis.

The research has compared with different classification algorithm included support vector machine, dynamic bayesian networks, artificial neural network and k-nearest neighbor classification algorithm [6]. Tom M. Mitchellconsider that the support vector machine has better learning and generalization ability among the small-sample, it can be well applied in medical image feature classification.Combined with the genetic algorithm and support vector machine (GA-SVM) and use it applied to the detection of lung nodules, use the genetic algorithm to optimize the support vector machinesmodelandachieved better results[7]. The particle swarm optimization algorithm has some advantages. After reading extensive literaturethere was rarely reported for the identification of lung nodules by PET / CT imaging features. Thisstudy using particle swarm optimization algorithm to optimize the parameters of support vector machine, then get a suitable classification model, and ultimately help doctors differentiating benign and malignant pulmonary nodules with higher accuracy.

EXPERIMENTAL SECTION

Feature Extraction: The studycollected 98 solitary pulmonary nodule cases since 2009. There are 54 cases are malignant pulmonary nodules and 44 cases are benign pulmonary nodules. Since the imageology character of pulmonary nodules is too complexfor the non-professional workers to recognize, this research asked two experienced chest radiologists. These two have worked in relevant area for more than 10 years. With their help,this research extracted 12 CT features: nodule shape, degree oflobulation, boundary definition, edge smoothness, the number of burr, density uniformity,airnodules,ground-glass opacity , calcification, vessel convergence , pleural indentation and the size of nodule.Each feature quantificationgradability from 0 to 10 by the experienced chest radiologists. This research view vacuole sign and inflatable bronchial sign as air nodules. For the size of the nodules quantification, the study selected the biggest level of the pulmonary nodules measuring its longest diameter and the shortest diameter, then get the average diameter.

Standard uptake value is used to describe the PET image character, calculation of standardised uptake value, is a common way to compare tumors patient to patient. Analyzed the PET / CT fusion images independently by two radiologists in PET / CT workstation.Measuring the nodules ROI maximum standard uptake value.

Super Vector Machine: The classification problem can be restricted to consideration of the two-type problem without loss of generality. In this problem the goal is to separate the two classes by a function which is induced from available examples. The goal is to produce a classifier that will work well on unseen examples, and it generalizes well. Consider the example in Fig1. Here there are many possible linear classifiers that can separate the data,but there is only green one that maximizes the margin. This linear classifier is termed the optimal separating hyperplane. Support Vector Machines (SVM) [8] is a relatively new technique introduced by Vapnik in the 90's. SVM is a mathematical technique for solving the classification problem described above. The principle of SVM is structural risk

Fig.1:Optimal Separating Hyperplane

minimization, taking bothtraining error and testing error minimized into account. Most importantly,you can use a limited number of training samples on SVM and get theincorrectresult as less as possible. At the same time, it is able to ensure the incorrectresult of independent test set as less as possible.Thus, it is a optimal generic learning machines.In thisresearch, the problem of the classification of benign and malignant pulmonary nodules, is a typical two-types classification problem and nonlinear separable, the original problem is expressed as follows:

$$\min_{\omega,\ b,\ \varepsilon} \frac{1}{2}\left|\left|\omega\right|\right|^2 + C\sum_{i=1}^{n} \varepsilon_i \ (1)$$

constraint condition：

$$y_i\big((\omega \cdot x_i) + b\big) \geq 1 - \varepsilon_i \varepsilon_i \geq \ 0 \ i = 1,2,\cdots,n$$

dual problem：

$$\min_{\mu} \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} y_i y_j \mu_i \mu_j \ K\big(x_i, x_j\big) - \sum_{j=1}^{n} \mu_j \ (2)$$

constraint condition：

$$\sum_{i=1}^{n} y_i \mu_i = 0 \;\; 0 \le \mu_i \le C \; i = 1,2,\cdots,n$$

Support vector machine for low-dimensional feature space do not separate problem can be converted to high-dimensional linear separable feature space, which requires the introduction of a kernel function K (x_i, x_j) to make nonlinear classification can be converted to the appropriate linear classification. The kernel function has a regularization parameter C, also referred to as a penalty factor which controls the balance between the two types. Select appropriate kernel function and parameter C solving equation (2), and the resulting decision function as the following formula:

$$f(x) = \mathrm{sgn}(\sum_{i=1}^{n} \mu_i^* y_i K(x_i, x_j) + b^*)$$

Different kernel function will lead to different support vector calculation method, widely used kernel functions: polynomial kernel function, linear kernel function, RBF kernel function, sigmoid kernel function. R Radial basis function (RBF) has better learning ability in different situations and apply for the small sample size, low-dimensional, so the study choose radial basis function as the kernel function of SVM. Radial basis function is expressed as follows:

$$K(x,y) = \exp\left\{-\frac{|x-y|^2}{\sigma^2}\right\}$$

This research need to determine the penalty factor C and RBF kernel function parameter σ in advance. Different C and σ values will lead to different classification accuracy, and the impact is huge. People selected the parameters randomly in the past; this will cause the results uncertainly. Therefore it is necessary to use some sort of algorithm to find the optimal parameter C and σ values.

Particle Swarm Optimization: In recent years, swarm intelligence algorithm gradually been applied to the kernel parameter optimization [9, 10]. Particle Swarm Optimization algorithm is a relatively new intelligent algorithm after genetic algorithm and ant colony algorithm. The PSO algorithm originated from a flock of birds foraging behavior. A birdrepresentsaparticle.A flock of birdsrepresentssolvable groups. The evolution of solvable groups is equivalent to the birdsmigration to another place. The good result of looking for the food whichequivalent tothe optimal solution in each generation evolution. Looking for a food source equivalent to the global optimalsolution [11,12].

The basic principles of particle swarm optimization algorithm. Assuming in the N-dimensional search space ， a community by Mparticles expressed as X, X=(X_1, X_2, ⋯, X_m), the i-th particle expressed as a N-dimensional vectorXi=（x_{i1}，x_{i2}，⋯，x_{in}）^{T}, the i-th particle flight speed is a N-dimensional vector V_i =（V_{i1}，V_{i2}，⋯，V_{in}）^{T}, P_i=(P_{i1}，P_{i2}，⋯，P_{in})^{T}on behalf of the individual extreme value , particle swarm entirely search for the optimal position, and P_b=（P_{b1}，P_{b2}，⋯，P_{bn}）^{T}on behalf ofthe global optimum. In an iterative process, particle updates their own speed and position through the individual extreme value and global extreme value. It says into the following formula:

$$V_{in}^{k+1} = \omega V_{in}^{k} + c_1 r_1 (P_{in}^{k} - X_{in}^{k}) + c_2 r_2 (P_{bn}^{k} - X_{in}^{k}) \quad (3)$$

$$X_{in}^{k+1} = X_{in}^{k} + V_{in}^{k+1} \quad (4)$$

Where ω is the inertia weight; n = 1,2, ., N; i = 1,2, ., M; k is the current iteration number; V_{in} is the velocity of the particle; c_1 and c_2, a non-negative constant as acceleration factor; r_1 and r_2,is random numberdistributed in interval [0, 1]. N-dimensional ($1 \le n \le N$) position change is range of [-X_{maxn}, X_{maxn}], the velocity change is range of [-V_{maxn}, V_{maxn}], within the search space of the problem can set $V_{maxn} = k * X_{maxn}$, $0.1 \le k \le 1$. If the position or velocity of the particles is pass over boundaries in iteration process, use the boundary value.

Particle Swarm Algorithm Optimizing SVM Parameters

As constituting a rule space, first need to discrete the parameter C and σ, and the parameter has a certain top and bottom limitation. When M particles placed in a suitable location of the space, particles will large-scale move, and leave pheromone at the path when over there, according to these pheromones update rules and particle path selection rules,these particles will eventually focus on a point with the continuously moving, and then this point is the optimal parameters (C, σ). The specific steps as follows:

Step1: Determine the basic parameters that the number of particles of the particle swarm algorithm M, acceleration coefficients C_1, C_2, and set the parameter C and σtop and bottom limitation. Set the termination condition to stop optimize when the average error of SVMtraining meet the accuracy requirements.
Step2: Placed M particles in the parameter space randomly, the number of moves marked S is set to 0.
Step3:The representative of the m-th particle parameter set is denoted by (C_m, σ_m), use the SVM to training samples.
Step4: Judge whether to meet the optimization termination conditions, meet then turn Step7 .otherwise, whether or not S is 0 then transferred Step5 otherwise turn step6.
Step5: According to the probability of path selection to move the position of the particle swarm, turn Step3.
Step6: Update the pheromone of the path, turn Step5.
Step7: Meet the requirements, optimization terminate and output optimal parameters

Operation flow chartshows inFig.2:

Fig.2:The flow chart of optimizing support vector machine parameters by particle swarm optimization

RESULTS AND DISCUSSION

Choose the support vector machine data set in this researchshows inTable1:

Table 1.The benign and malignant pulmonary nodules sample size used by the support vector machine

	Training data	Testing data	Total
Malignancy	30	24	54
Benignancy	26	18	44
Total	56	42	98

The particle swarm algorithm to optimize the parameters of support vector machine. Must initialized the parameter of particle swarm in advance. The study set the M is 20, acceleration coefficient $C_1 = C_2 = 2$, $\omega = 1$, the number of iterations is 50,set σ range in [0,1], the penalty parameter C range in [0,10].

The support vector machine through particle swarm optimizationalgorithm obtain optimization parameters C = 6.3042, σ = 0.4123, classification of benign and malignant solitary pulmonary nodules results obtained, as shown in Table 2.

Table 2.The classification results of benign and malignant pulmonary nodules by two kinds of support vector machine

	Malignancy	Benignancy	Average accuracy/%
Support vector machine0	18/24	13/18	73.6
Support vector machine1	21/24	16/18	88.1

SVM0 isaclassification model which choosesparametersrandomly. Testing the modeluse test samples based on SVM0. The result shows 18 cases right in24 malignant nodules, and 13 cases right in 18 benign nodules. Accuracy rate is 0.75 and 0.72 respectively. The average accuracy rate is 0.736. SVM1 isaclassification modelwhichuses the optimized parameters. The study get 21 cases right in 24malignantnodules, and 16 cases right in 18 benign nodules by SVM1. Accuracy rate is 0.87 and 0.88respectively.The average accuracy rate is 0.879.Therefore, it can be seen that the classification model by parameter optimization compared to the classification model without parameter optimization. The accuracy rate was significantly improved on the identification of lung nodules classification of benign and malignant.

The study compared the results generated through the two models.Althoughthe number of sample set is low, good classification results still can be obtained.It is shows that using particle swarm optimization on support vector machine parametric search can get an appropriate SVM classification model.thus the solitary pulmonary nodules of benign and malignant can get abetter classification result.

CONCLUSION

The contribution of thisresearch is propose a novel classification model method based on particle swarm optimization combined support vector machines by pulmonary nodules PET/CT imaging features hierarchical quantification, and ultimately obtain the support vector machine model which have higher classification accuracy. When doctors reading the PET/CT images, they only choose a little features by influence of experience, they can’t use the characteristics comprehensive and effective that affect the accuracy of diagnosis. However, optimized support vector machine (PSO-SVM) can consistently integrated CT and PET image information to summarized that comprehensive ability is superior to the general classification model, also can provide theoretical references to doctors especially inexperienced young doctors .

Acknowledgements
Thanks the fund of National Natural Science Foundation of China (61202163, 61373100); Natural Science Foundation of Shanxi Province(2012011015-1); Scientific and Technological Project in Shanxi Province (20120313032-3).

REFERENCES

[1] Soubani A O. *Postgraduate medical journal*, **2008**, 84(995): 459-466
[2] Van Rens M T M, de la RiviereA B, Elbers H R J, et al. *CHEST Journal*, **2000**, 117(2): 374-379.
[3] Roy A S, Armato III S G, Wilson A, et al. *Medical physics*, **2006**, 33: 1133.
[4] Varoli F, Vergani C, Caminiti R, et al. *European journal of cardio-thoracic surgery*, **2008**, 33(3): 461-465.
[5] Lonsdale M N, Beyer T. *European journal of radiology*, **2010**, 73(3): 452-460.
[6] Mitchell T M, Hutchinson R, Niculescu R S, et al. *Machine Learning*, **2004**, 57(1-2): 145-175.
[7] Jaffar M A, Hussain A, Jabeen F, et al. GA-SVM Based Lungs Nodule Detection and Classification[M]//Signal Processing, Image Processing and Pattern Recognition. Springer Berlin Heidelberg, **2009**: 133-140.
[8] Vapnik V. The nature of statistical learning theory[M]. springer, **2000**.
[9] Dautenhahn K. Book review: Swarm intelligence by jameskennedy, russell c. eberhart, with yuhuishi[J]. *Genetic Programming and Evolvable Machines*, **2002**, 3(1): 93-97.
[10] J Kennedy;REberhart; Proceedings of the International Conference on Neural Networks , **1995**, 1942-1948 .
[11] A Ahmadian; AMostafa; MD Abolhassani; NRiahiAlam;Proceedings of the 7th International Conference on Signal Processing, August 31-September 4, **2004**, IEEE Xplore London, 971-974.
[12] Eberhart R C, Shi Y, Kennedy J. Swarm intelligence[M]. Access Online via Elsevier, **2001**.