SUPPORT VECTOR REGRESSION FILTERING FOR REDUCTION OF FALSE POSITIVES IN A MASS DETECTION CAD SCHEME: PRELIMINARY RESULTS

Enrico Angelini, Renato Campanini and Alessandro Riccardi

Department of Physics, University of Bologna, Italy
Corresponding author: riccardi@bo.infn.it

ABSTRACT

Reduction of False Positive signals (FPR) is a fundamental, yet awkward, step in computer aided mass detection schemes. This paper describes preliminary results of a filtering approach to FPR based on Support Vector Regression (SVR), a machine learning algorithm arising from a well-founded theoretical framework, the Statistical Learning Theory, which has recently proved to be superior to the conventional Neural Network framework for both classification and regression tasks: indeed, the proposed filtering method belongs to the family of neural filters.

The SVR filter is forced to associate subregions extracted from input images, masses and non-masses, to continuous output values ranging from 0 to 1 representing a measure of the presence in the subregion of a mass. A weighted sum of outputs over each image is used to accomplish the FPR task. In the test phase, this approach reached promising results, retaining 87% of masses while reducing False Positives to 62%.

SUPPORT VECTOR REGRESSION FILTERING FOR REDUCTION OF FALSE POSITIVES IN A MASS DETECTION CAD SCHEME: PRELIMINARY RESULTS

Enrico Angelini, Renato Campanini and Alessandro Riccardi

Department of Physics, University of Bologna, Italy
Corresponding author: riccardi@bo.infn.it

ABSTRACT

Reduction of False Positive signals (FPR) is a fundamental, yet awkward, step in computer aided mass detection schemes. This paper describes preliminary results of a filtering approach to FPR based on Support Vector Regression (SVR), a machine learning algorithm arising from a well-founded theoretical framework, the Statistical Learning Theory, which has recently proved to be superior to the conventional Neural Network framework for both classification and regression tasks: indeed, the proposed filtering method belongs to the family of neural filters.

The SVR filter is forced to associate subregions extracted from input images, masses and non-masses, to continuous output values ranging from 0 to 1 representing a measure of the presence in the subregion of a mass. A weighted sum of outputs over each image is used to accomplish the FPR task. In the test phase, this approach reached promising results, retaining 87% of masses while reducing False Positives to 62%.

1. INTRODUCTION

Breast cancer is the most widespread form of cancer among women in many countries. Computer Aided Detection (CAD) schemes have been developed by many authors over the years to help radiologists to improve the detection rate of cancer lesions - microcalcifications and masses - in mammograms [1][2][3]. In particular, mass detection has been shown to be extremely difficult, due to the great variety in shape, size, texture and subtlety. An essential step in CAD systems is False Positive Reduction (FPR): elimination of signals erroneously detected by the system.

In this paper we introduce a Support Vector Regression filtering approach to FPR in an automated mass detection system. The proposed approach consist of two steps: the first one belongs to the family of neural filters [4][5], but is based on Support Vector Machine (SVM), a class of learning algorithm which has proved in recent years to be superior to the conventional Neural Network method for both classification and regression tasks [6][7], hence its application to neural-like image processing looks very appealing. The second step is a simple way to take into account information given by SVR filter, in order to decide whether the analyzed signal is a False Positive (FP) or not.

2. MATERIALS

2.1. The mass detection system

The CAD system we have applied FPR to is described in [3]. Essentially, it considers mass detection in mammograms as a two-class pattern recognition problem. The detection procedure is the following: firstly the segmented mammogram is scanned with smaller overlapping windows of various sizes, secondly the information in each of these crops is codified by means of an overcomplete wavelet representation, and finally this great amount of information is given to a Support Vector Machine (SVM) classifier, which is devoted to separate crops with masses from those without. All crops, taken at various sizes, are resampled to be 64 pixels in side, before being codified through wavelet transform and passed to the classification step. The SVM classifier is trained with mass crops (namely, ground truth crops), whose size and position are given by radiologists, and non-mass crops, taken at random from within the breast area. The system at this stage analyses approximately 10^5 crops per image on average, and about $0.05 \div 0.1\%$ of them are approximately classified as Positive, that is, contain a mass. Most of them are simply FPs, and need to be eliminated as much as possible, a task that is accomplished via a cascaded SVM classifier.

The complete CAD scheme comprises a final step, which consists of an ensemble of experts judging candidates from at least three systems as that described above, only different in some parameters. Anyway, here we consider the SVR filtering approach to FPR as an alternative for the cascaded classifier in a single system.

2.2. The database of mammograms

The data set used consists of 69 mammograms with dimensions 2816 x 2048, pixel size 85 μm and bitdepth 13, contains 69 ground truth regions and was acquired with the Full Field Digital Mammography apparatus (FFDM) *Giotto Image MD* (IMS srl, Bologna, Italy). First stage of CAD processing (described in section 2.1) of this data set gives about 8000 crops classified as masses, of which only about 900 can be considered containing a true mass, the rest being False Positives. This high number of Positive Crops is due to the fact that at this stage of analysis redundant crops, that is, showing the same signal, are not yet clustered together (see [3]). This redundancy comes from the use of many different window sizes (from 8 to 40 mm in side) to scan the mammogram. For the same reason, the 69 ground truth regions give rise to 349 ground truth crops.

3. SUPPORT VECTOR REGRESSION

Support Vector Machine is a class of supervised learning algorithms for nonlinear classification and regression, deep-rooted in the Statistical Learning Theory [6], and at present considered the state-of-art in the machine learning field: indeed, it exhibits many useful properties that make it a suitable candidate for neural filtering. Among these properties are: 1) the training process is made by solving a quadratic optimization problem, hence no local minima trapping is ever encountered; 2) the architecture of the machine is automatically set by the solution of the optimization problem, which makes the training be quite easy, and, moreover, allows for a very low risk of poor generalization. These properties guarantee in general a learning procedure simpler and faster than that conveyed with a conventional multilayer Artificial Neural Network, or with an RBF.

ν-Support Vector Regression (ν-SVR) [8] is the SVM algorithm for regression estimation considered here. Given the vector data set $(\mathbf{x}_i, y_i)_{i=1}^{l}$, in the linear case this algorithm seeks to estimate

$$f(\mathbf{x}) = (\mathbf{w} \cdot \mathbf{x}) + b, \ \ \mathbf{w}, \mathbf{x} \in R^M, b \in R \quad (1)$$

by making use of the so-called ε-insensitive loss function:

$$|y - f(\mathbf{x})|_\varepsilon = \max[0, |y - f(\mathbf{x})| - \varepsilon]. \quad (2)$$

The algorithm consists in the solution of the following constrained quadratic optimization problem:

minimize

$$\tau(\mathbf{w}, \xi^{(*)}, \varepsilon) = \frac{1}{2}\|\mathbf{w}\|^2 + C\left(\nu\varepsilon + \frac{1}{l}\sum_{i=1}^{l}(\xi_i + \xi_i^*)\right) \quad (3)$$

subject to

$$\begin{aligned} &((\mathbf{w} \cdot \mathbf{x}_i) + b) - y_i \le \varepsilon + \xi_i \\ &y_i - ((\mathbf{w} \cdot \mathbf{x}_i) + b) \le \varepsilon + \xi_i^* \\ &\xi_i^* \ge 0, \varepsilon \ge 0. \end{aligned} \quad (4)$$

At each point $\mathbf{x}_i$ the maximum allowed error is ε: every higher error goes into slack variables $\xi_i^{(*)}$ and is penalized trough constant C, chosen a priori. The other parameter to choose a priori is ν, which is bounded between 0 and 1 and governs the trade off between good fitting of data and smoothness of solution, which is directly linked to generalization ability. The usual way to solve the SVM optimization problem is to transform its Lagrangian formulation into the corresponding Wolfe dual form [6]. The regression estimate, which depends via the Lagrange multipliers α and α^* on a subset of training vectors called Support Vectors, is:

$$f(\mathbf{x}) = \sum_{i=1}^{l}(\alpha_i^* - \alpha_i)(\mathbf{x} \cdot \mathbf{x}_i) + b \quad (5)$$

The general case of non-linear regression can be very easily handled by substituting a kernel $k(\mathbf{x}_i,\mathbf{x}_j)$ to the dot product in the dual formulation and consequently in (5). It is worth noting that kernel functions, which correspond to dot product in a feature space given by a nonlinear transformation ϕ of the data vectors in the input space,

$$k(\mathbf{x}_i, \mathbf{x}_j) = (\phi(\mathbf{x}_i) \cdot \phi(\mathbf{x}_j)) \quad (6)$$

show a large variety of forms: polynomial, RBF, sparse polynomial, neural network, string kernel, etc [7].

4. FPR VIA SVR FILTERING

4.1. SVR filtering

The SVR learning algorithm acts as a filter because it is able to associate to each input crop an output image, which is subsequently used to determine if the crop contains or not a mass.

The algorithm is trained with a set of positive and negative samples. The positive ones are associated to a teacher image containing a 2 dimensional Gaussian function, whilst the negatives are associated to an image of zeros. Overlapping subcrops are extracted from each crop by scanning it with a square window of size R_x, along rows and columns, by moving the center of R_x by a certain amount of pixels P at a time, without letting the window go out of the crop. At the same time, the same window scans the teacher image, and the center pixel value of this subimage is retained (see Fig. 1). Subcrops and their coupled teacher image values are then used for SVR training, in which the regression algorithm is forced to associate the corresponding value *y* to every given crop pattern ***x*** of length R_x^2. The 2 dimensional Gaussian teacher image values range from 0 to 1, and can be considered as a measure of "how much" of a mass is enclosed by the scanning window: indeed, the mass is positioned in the middle of each training crop, covering about 20÷25% of crop area, and the Gaussian as well is centered in the middle of the teacher image (see Fig. 1). The number of subcrops extracted from each crop, depending on the size of R_x and on the value of P, increase when R_x or P decrease.

During test phase, from each scanned crop an output image is created: every pixel in it contains the “opinion” of SVR about the presence of a mass, or a part of it, in the corresponding analyzed subcrop.

4.2. Scoring and thresholding

To obtain a single measure $M(c_i)$ of the presence of a mass in the crop c_i, the output image values $\boldsymbol{y}^i$ are put together using a Gaussian weighted sum, so that higher measures can be achieved by positive crops whose outputs are in good agreement with the Gaussian teacher image (high values in the middle getting lower towards borders):

$$M(c_i) = \sum y^i_{n,m} w(\sigma_g)_{n,m}. \qquad (7)$$

The weight $\boldsymbol{w}(\sigma_g)$ represents the Gaussian image with sigma σ_g.

Typically, the distribution of measures M obtained from a group of positive and negative crops overlap, but shows two peaks: the rightmost is related to mass crops, the other to non-mass crops. By putting a threshold on M it is possible to distinguish, to a certain degree, the two kinds of crops. The optimal value of the threshold is determined during the training phase.

5. RESULTS

All a-priori parameters have been chosen through a Cross–Validation (CV) procedure, and test phase has been performed afterward on different independent sets. Training set was given 26 ground truth crops and a maximum of 104 randomly chosen non-mass crops.

5.1. Cross-Validation training

We opted for CV because of the low number of ground truth crops at disposal: in particular we performed M-Fold CV with M = 13, putting 2 positive crops in each fold. The number of negatives varied from 26 to 104, which gave us the best performances. Due to the high demand in computational resources we decided to train each fold alone, instead than M - 1 together as in usual M-Fold CV, and then calculate the mean result of the M - 1 trained SVRs over the M^{th}, iteratively for the M folds, to obtain the CV results: this procedure belongs to the ensemble methods of machine learning [9].

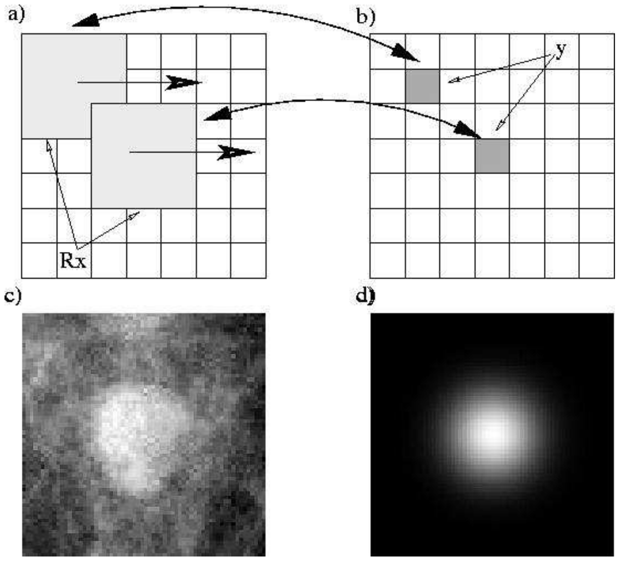

Fig. 1. The action of the filter over the analyzed crop: a) as the R_x window scans the crop, creating subcrops, b) a value y is associated to each subcrop: this value y is drawn from the teacher image in the training phase, and is produced by SVR in the test phase. In c) it is shown a positive sample from the training set: the mass is clearly visible in the center of the crop. In d) there is depicted, as a gray level image, the 2 dimensional Gaussian function used in the training phase as teacher image for positive crops.

We used SVR implementation in Spider package (http://www.kyb.tuebingen.mpg.de/bs/people/spider/).

We obtained the best performances with these parameters: R_x 17 pixels in side, resulting in 17^2 = 289 dimensions for vector **x**, P equal to 4, resulting in a gain of only 144 vectors per crop, the teacher image containing the Gaussian function until ±4 sigma. In addition, SVR parameters were: RBF kernel with sigma equal to 3, C = 1000, and ν = 0.4, but varying almost in its entire range without affecting the results. As usual in SVM algorithms, data vectors underwent a whitening procedure [10]. Training time was 260 sec on average for each fold on a 2.6 GHz Pentium Xeon with 2Gb RAM running Linux.

The FPR procedure obtained very good CV results: it could be able to reduce FPs to 65±10% and 36±10% while retaining at the same time 95±5% and 88±12% of Positive crops, at two different threshold values on M, respectively. Uncertainties over these results were calculated with a 95% binomial confidence interval estimation [11].

5.1. Test

Test was performed on crops coming from the 43 images (with 43 ground truth regions) that were not used in the CV training phase.

The first test set consisted of 205 ground truth crops, extracted at all the scales used by the system, and 510 non-mass crops randomly chosen from the FPs given in output by the CAD after scanning the 43 images. For the same threshold values considered before, 94±3% and 81±5% of mass crops and 82±3% and 62±4% of FP crops were retained. It is worth noting that the positive crops considered here are different from those used in training, showing in general a very different mass-area/crop-area ratio, because of the many scales they were acquired at. Anyway, the sensitivity over positives is almost the same as that found in CV, which means this ratio is not a critical parameter of the approach. On the other side the percentage of FPs does not drop as quick as expected from CV results. The main cause of this fact should be found in the low number of training samples: the negative ones present a great variability, which has probably not been captured by the small training set of only 104 randomly chosen negatives; besides, also the 26 positives considered for training could hardly be considered representative of their class, which anyway shows less variability. The optimal number of training samples is an important, yet open, question: more data are necessary to obtain an answer.

The second test set consisted of the same FPs and of 546 crops classified as containing a mass by the CAD scheme: these are the usual data that should undergo FPR. At the same values of FPs as before, 96±2% and 87±3% of mass crops survived the thresholding phase. These percentages agree with the previous ones.

6. CONCLUSIONS AND FUTURE WORK

The proposed two-step method reached promising results as an FPR stage in a CAD scheme for mass detection, being able to eliminate about 38% of FPs at a cost of only 13% of mass crops. This rate of reduction of FP signals is at present no better than that used in the CAD scheme in [3], and further work needs to be done. Anyway, the method has proved to be valuable: in particular its core, SVR filtering, has showed to be an effective and easy-to-train way to extract the information concerning the kind of tissue in the crop, and should be pursued further. The second and final step, on the other side, is the one that needs more improvements: indeed, instead of weighted scoring, a more advanced method able to effectively deal with the information carried by the filtered images - such as a multidimensional classifier like SVM - should be taken into account. In addition, the question of the optimal number of samples in training is a relevant issue, as it is in general for machine learning based approaches, and should also be tackled.

Future work should involve: 1) the assessment of the performances of the proposed approach at varying number of training samples, and 2) the devising of more effective methods to deal with SVR filtering outputs.

7. REFERENCES

[1] H. P. Chan, K. Doi, S. Galhotra, C. J. Vyborny, H. MacMahon and P. M. Jokich, “Image feature analysis and computer aided diagnosis in digital radiography: automated detection of microcalcifications in mammography”, *Med. Phys.*, 14, pp. 538-548, 1987.

[2] N. Petrick, B. Sahiner, H. P. Chan, M. A. Helvie, S. Paquerault and L. M. Hadjiiski, “Breast cancer detection: evaluation of a mass-detection algorithm for computer aided diagnosis – experience in 263 patients”, *Radiology*, 224, pp. 217-224, 2002.

[3] R. Campanini, D. Dongiovanni, E. Iampieri, N. Lanconelli, M. Masotti, G. Palermo, A. Riccardi and M. Roffilli, “A novel featureless approach to mass detection in digital mammograms based on support vector machines”, *Physics in Medicine and Biology*, IOP Publishing, UK, 49, pp. 961-975, 2004.

[4] K. Suzuki, I. Horiba and N. Sugie, “Neural Edge Enhancer for Supervised Edge Enhancement from Noisy Images”, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 25, no. 12, pp. 1582-1596, December 2003.

[5] K. Suzuki, S. G. Armato, F. Li, S. Sone and K. Doi, “Massive training artificial neural network (MTANN) for reduction of false positives in computerized detection of lung nodules in low-dose computed tomography”, *Med. Phys.*, 30(7), pp. 1602-1617, July 2003.

[6] Vapnik, V., *The Nature of Statistical Learning Theory*, Springer Verlag, New York, 1999.

[7] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machine*, Cambridge University Press, 2000.

[8] B. Schoelkopf, A. Smola, R. Williamson, and P. Bartlett, “New support vector algorithms”, *Neural Computation*, 12(5), pp.1207-1245, May 2000.

[9] T. G. Dietterich, “Ensemble methods in machine learning”, *Lecture Notes in Computer Science,* Vol 1857, pp. 1-15, Springer-Verlag, 2000.

[10] R. O. Duda, P. E. Hart and D. G. Stork, *Pattern Classification (2nd Edition)*, Wiley Interscience, 2000.

[11] L. D. Brown, T. T. Cai and A. DasGupta, “Confidence intervals for a binomial proportion and asymptotic expansions ”, *The Annals of Statistics*, Vol 30, No 1, pp. 160-201, 2002.