Channelized Relevance Vector Machine as a Numerical Observer for Cardiac Perfusion Defect Detection Task

Mahdi M. Kalayeh[a], Thibault Marin[a], P. Hendrik Pretorius[b], Miles N. Wernick[a], Yongyi Yang[a] and Jovan G. Brankov[a]

[a]Medical Imaging Research Center (MIRC), Department of Electrical and Computer Engineering, Illinois Institute of Technology, 3440 S. Dearborn St., Chicago, IL, 60616;

[b]Department of Radiology, Division of Nuclear Medicine, University of Massachusetts Medical School, 55 Lake Ave N, Worcester, MA, 01655, USA

ABSTRACT

In this paper, we present a numerical observer for image quality assessment, aiming to predict human observer accuracy in a cardiac perfusion defect detection task for single-photon emission computed tomography (SPECT). In medical imaging, image quality should be assessed by evaluating the human observer accuracy for a specific diagnostic task. This approach is known as task-based assessment. Such evaluations are important for optimizing and testing imaging devices and algorithms. Unfortunately, human observer studies with expert readers are costly and time-demanding. To address this problem, numerical observers have been developed as a surrogate for human readers to predict human diagnostic performance. The channelized Hotelling observer (CHO) with internal noise model has been found to predict human performance well in some situations, but does not always generalize well to unseen data. We have argued in the past that finding a model to predict human observers could be viewed as a machine learning problem. Following this approach, in this paper we propose a channelized relevance vector machine (CRVM) to predict human diagnostic scores in a detection task. We have previously used channelized support vector machines (CSVM) to predict human scores and have shown that this approach offers better and more robust predictions than the classical CHO method. The comparison of the proposed CRVM with our previously introduced CSVM method suggests that CRVM can achieve similar generalization accuracy, while dramatically reducing model complexity and computation time.

Keywords: Numerical observer, Bayesian learning, Relevance vector machine, RBF Kernels, Perfusion defect detection, SPECT, AUC, MSE

1. INTRODUCTION

Diagnostic accuracy is the ultimate criterion for quantifying image quality; however, human observer (HO) studies are costly and time-consuming. Therefore, numerical observers have received growing attention as a practical alternative for predicting human diagnostic performance and facilitating automated image quality assessment. For defect detection tasks, the channelized Hotelling observer (CHO)[1], introduces a statistical model based on features extracted using bandpass filters and a linear discriminant. This method, which is commonly used as a surrogate for human observers in a perfusion defect detection task, gained popularity due to its good performance in predicting human observer's performance in some settings.[2] However, in practice, CHO detection performance is not always well correlated to human observers. One solution is to integrate a so-called internal noise term in the model.[3]

In a previous study[4], we rephrased the numerical observer as a machine learning problem, and used support vector machines (SVM)[5] to predict human observers' scores. It resulted in a channelized support vector machine (CSVM) as a numerical observer to predict the human observer's performance in a perfusion defect detection task. SVM is based on statistical learning theory[5], utilizing the so-called structural risk minimization (SRM) principle, and yields models that generalize well to data outside of the training set. This method outperformed the CHO in our prior studies.

Medical Imaging 2011: Image Perception, Observer Performance, and Technology Assessment, edited by David J. Manning, Craig K. Abbey, Proc. of SPIE Vol. 7966, 79660E · © 2011 SPIE
CCC code: 1605-7422/11/$18 · doi: 10.1117/12.878176

In this paper, we consider an alternative method that replaces the SVM with a relevance vector machine (RVM)[6], a modern Bayesian learning approach that has several important benefits over SVM: it uses dramatically fewer kernel functions, has lower model complexity, results in greatly reduced computational cost, and can use arbitrary kernels (e.g. non-'Mercer' kernels)[6].

2. METHODOLOGY

In this section, we describe the proposed RVM-based method for prediction of the human observer scores and the training-testing paradigm.

2.1 Relevance vector machine

The relevance vector machine (RVM)[6,7] is a supervised Bayesian learning algorithm which can be used to predict human scores, $\{t_n\}_{n=1}^{N}$, from extracted image features $\{\mathbf{x}_n\}_{n=1}^{N}$, where N is the number of training samples (the training procedure will be explained comprehensively in Sec. 2.2). In our case these features are extracted from the images as described later in Sec. 3.3. The relationship between features and scores in the RVM (regression model) has the following form:

$$t = y(\mathbf{x}) + \varepsilon \,, \tag{1}$$

where $\varepsilon : N(0,\sigma^2)$ is the additive noise, t denotes the target score, $y(\mathbf{x})$ is the regression model predicting the human score and $\mathbf{x}$ is the set of extracted channelized features from a particular image. We can rewrite the model as Eq. 2 where $\{w_n\}_{n=0}^{N}$ model weights and $K(.,.)$ represents the kernel function that maps the feature space to a non-linear space where a linear regression is applied.

$$t = \sum_{n=1}^{N} w_n K(\mathbf{x}, \mathbf{x}_n) + w_0 + \varepsilon, \tag{2}$$

To palliate the high computational cost of a non linear regression in feature space, we use kernel functions to transfer from feature space to a higher dimensional space. The linear regression can then be applied in the augmented space. Thus, the model is built as a weighted linear combination of the kernel function of feature values. In our case, we used the radial basis function (RBF) which is formulized as Eq. 3 and parameterized by γ_K, determining the kernel width:

$$K(\mathbf{x},\mathbf{z}) = \exp\left(-\frac{\|\mathbf{x}-\mathbf{z}\|^2}{2\gamma_K^2}\right). \tag{3}$$

RVM uses a hierarchical prior probability distribution on the weights $\{w_n\}$ encouraging sparsity (mostly zero weights), similar to compressive sensing. The prior is defined as:

$$P(\mathbf{w}|\alpha) = \prod_{n=1}^{N} N(0, \alpha_n^{-1}), \tag{4}$$

where $\mathbf{w} = [w_0, w_1, \ldots, w_N]^T$ and $\mathbf{á} = [\alpha_0, \alpha_1, ..., \alpha_N]^T$ determine the precision of each particular weight and $N(.,.)$ denotes the Gaussian probability function. The RVM methodology aims at maximizing the posterior probability:

$$P(\mathbf{w}, \alpha, \sigma^2 \,|\, \mathbf{t}) = P(\mathbf{w} \,|\, \mathbf{t}, \alpha, \sigma^2) \times P(\alpha, \sigma^2 \,|\, \mathbf{t}), \tag{5}$$

where $\mathbf{t} = [t_1, t_2, \ldots, t_N]^T$. By assuming non-informative Gamma priors on hyper-parameters σ^2 and $\{\alpha_n\}_{n=0}^{N}$, one can show that:

$$\mathbf{w}^* = \frac{1}{\sigma^2}\left(\frac{1}{\sigma^2}\Phi^T\Phi + diag(\alpha)\right)^{-1}\Phi^T\mathbf{t}\,. \tag{6}$$

where $\mathbf{\Phi}$ is a design matrix with elements $\left[\mathbf{\Phi}\right]_{n,m} = K\left(\mathbf{x}_n, \mathbf{x}_{m-1}\right)$ and $\left[\mathbf{\Phi}\right]_{n,1} = 1$.We assume uniform (in logarithmic scale) distributions for hyper priors (α, σ^2) which results in a very pleasing consequence of scale invariance. Since all scales are equally likely, predictions are independent of linear scaling of both $\mathbf{t}$ and design matrix ($\mathbf{\Phi}$) outputs. The hyper-parameters cannot be obtained in a closed form so they are assigned some initial values and the final optimal value will be obtained using an iterative approach. In the iterative algorithm, we are aiming at estimating the mean, $\boldsymbol{\mu}$, and covariance, $\boldsymbol{\Sigma}$, of the weights at each iteration and using them to recalculate the new hyper-parameters. Eq. 7 to Eq. 10 show the process of updating the hyper-parameters at each iteration of the optimization.

$$\boldsymbol{\Sigma} = \left(diag\left(\alpha\right) + \frac{1}{\sigma^2}\mathbf{\Phi}^T\mathbf{\Phi}\right)^{-1}, \tag{7}$$

$$\boldsymbol{\mu} = \frac{1}{\sigma^2}\boldsymbol{\Sigma}\Phi^T\,\boldsymbol{t}, \tag{8}$$

$$\alpha_i^{new} = {1-\alpha_i\Sigma_{ii}}\Big/{\mu_i^2}\,, \tag{9}$$

$$\sigma^{2^{new}} = \frac{\left\|\mathbf{t} - \Phi\boldsymbol{\mu}\right\|}{N - \sum_{i=1}^{N}\left(1-\alpha_i\Sigma_{ii}\right)}. \tag{10}$$

At the end of the iterative procedure, only a few weights are non-zero and thus only a few number of samples are used for the regression model. These samples are called relevance vectors (RVs) which describe the learning model. The lower the number of basis functions (RVs), the simpler the prediction model.

2.2 Training and testing of CRVM

In this work, as mentioned previously, we used radial basis functions for $K(.,.)$, parameterized by kernel width γ_K. To ensure good generalization performance, the optimal width, γ_K, was found using 6-fold cross validation. The number of parameters incorporated in the K-fold cross validation determine the most significant portion of the computation cost. In the CSVM method, we had two more parameters in addition to the one we have for the CRVM. These were in charge of determining the level of penalty we would assign to make a trade-off between the training and testing error for realization of structural risk minimization (SRM)[5].

The higher number of design parameters results in a considerably larger number of possible combination of design parameters. Since the K-fold cross validation performs a grid search in the parameter space, reducing the number of parameters to select from significantly reduces the computation time required for parameter selection. A quantified comparison of the parameter selection computation between CSVM and CRVM will be reported in Table. 2. The selected set of design parameters (i.e. the optimal width γ_K) will be used for training of the numerical observer. Once trained, the numerical observer is tested on a dataset containing samples different from the ones used for training. The creation of training and testing dataset will be explained in Sec. 3.4.

3. RESULTS

In this section, we compare CRVM and CSVM generalization performance in terms of mean squared error (MSE) of score predictions and the area under the receiver operating characteristic (ROC) curve (AUC)[8,9], model complexity and computation cost. We used the MOSEK[10] optimization toolbox to implement quadratic programming used in SVM and SPIDER[11] machine learning environment to implement mathematical calculations of RVM. Performance comparison of CSVM and CHO can be found in Ref. 4.

3.1 Myocardial-Perfusion imaging

In this work we used data from a simulated myocardial-perfusion SPECT study previously published by Narayanan.[12] A Mathematical Cardiac Torso (MCAT) phantom[13] aiming at simulating respiratory motion, the wringing motion of beating heart and heart chamber contraction was used to model a perfusion defect at a fixed position in the left-ventricular wall (signal known exactly (SKE) paradigm). The MCAT phantom was used to generate images in a $128\times128\times128$ grid with a voxel size of 0.317 cm. To simulate a low-energy and high-resolution collimator, the effects of non uniform attenuation, photon scatter and distance-dependent resolution (DDR), the SIMIND software[14] was used to generate projection images. The projections were reconstructed using the ordered subsets expectation-maximization (OSEM) algorithm[15] with one and five effective iterations. Reconstruction was performed for one hundred noise realizations for each number of iteration.

Next, images were low-pass filtered by 3-D Gaussian filters with 0, 1, 2, 3, 4 and 5 pixels as full width at half maximum (FWHM). This resulted in six different level of smoothing for each particular iteration number. The final dataset contained 12 different sets of reconstructed images, each one corresponding to an FWHM and a number of effective iterations. In Figure 1, we show several examples of the OSEM-reconstructed images.

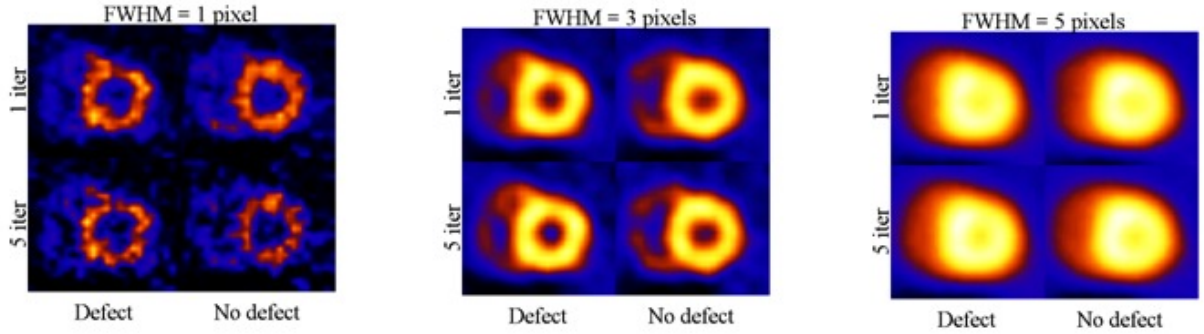

Figure 1. Three levels of smoothing for defect present and defect absent cases reconstructed with 1 and 5 iterations

3.2 Human observer study

In a human observer study, the observer's task was to determine the presence of a myocardium defect based on the intensity change in the image. Two observers were asked to score the defect presence likelihood on a 1 to 6 discrete scale in a signal known exactly (SKE) environment. This procedure was repeated over 100 images for each specific combination of FWHM and effective iteration numbers. The diagnostic accuracy of human observer (HO) is summarized by the area under receiver operating curve (AUC).[8]

3.3 Feature extraction

Myers and Barrett[1] demonstrated that applying non overlapping bandpass filters to images could be utilized to extract features used in a Hotelling observer to predict human diagnostic performance in a perfusion defect detection task. These channels are 2-D rotationally symmetric bandpass filters, which are intended to model the human visual system[1]. In our experiments, we used four bandpass filters with cutoff frequencies between 0.5, 0.25, 0.125, 0.0625 and 0.0313 cycles/pixel, resulting in four features per image.

Figure 2 shows four feature extracting channels in the frequency domain. Since we are working in a SKE environment, the four feature extracting channels are applied on a windowed area centered at the known location of simulated perfusion defect in the MCAT phantom.[12] Note that the window used in this work had a size of 51×51 pixels.

Figure 2 Four feature extracting channels in frequency domain

3.4 Numerical observer evaluation

An example application in which one might use a numerical observer is to optimize the parameters of an image-reconstruction algorithm. In such a situation, it is important that the numerical observer is able to predict human observer performance accurately even if it has not been exposed to images reconstructed using the specific method studied. To test such a situation, we created two sets of image data: In the first set, images were reconstructed with one iteration of the OSEM algorithm with varying levels of post-reconstruction smoothing; in the second set, the smoothing parameter was similarly varied, but five iterations of OSEM were used. In our evaluation study, we trained using images from one of these two sets, and tested on images from the other. Thus, the test set consisted of a different distribution of images from the training set. Such a situation is very typical in practice.

In Figure 3, we compare CSVM and CRVM in terms of the mean square error (MSE) between predicted and human scores, as well as the predicted AUC. Figure 3 shows that both CRVM and CSVM are able to closely follow the human AUC curve, but the MSE of predicted human scores is lower for CRVM than for CSVM. Furthermore, Table 1 indicates that the variance of the MSE is reduced by an order of magnitude when using the proposed CRVM compared to CSVM. Note that the object of the learning machine is to predict the human scores, but in the end the numerical observer is concerned principally with aggregate performance, as measured by the fidelity of the AUC values.

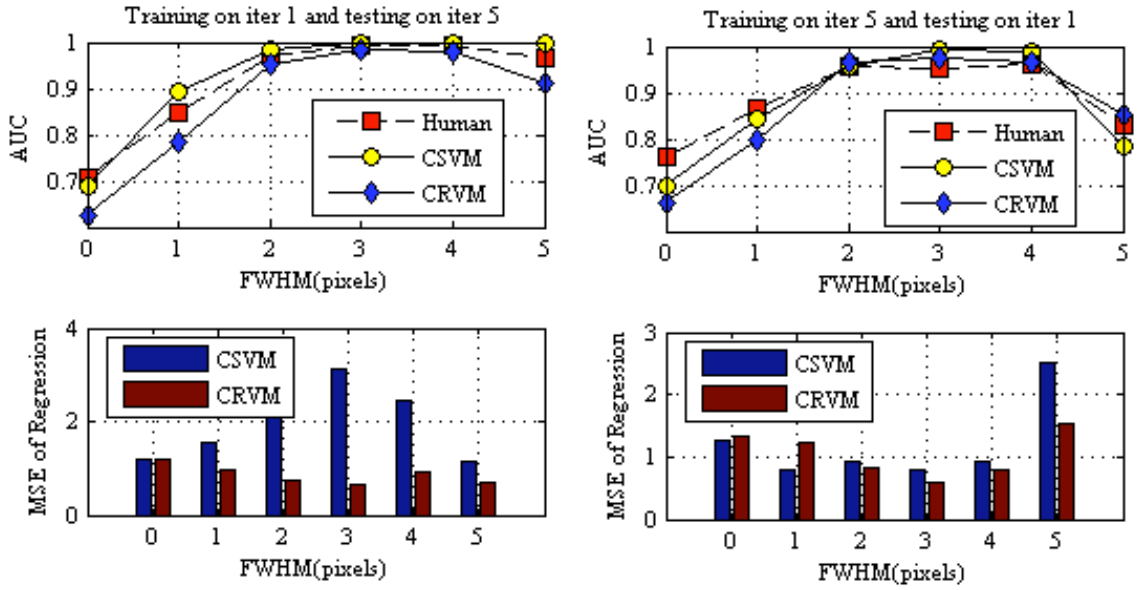

Figure 3. CRVM and CSVM performances on testing data set

Table 1 Generalization error for CSVM and CRVM

Training Dataset	Testing Dataset	Numerical Observer	Squared Mean	Variance
1 iter.	5 iter.	CSVM	1.9789	0.6561
1 iter.	5 iter.	CRVM	0.8576	0.0402
5 iter.	1 iter.	CSVM	1.2064	0.4477
5 iter.	1 iter.	CRVM	1.0474	0.1372

As shown in Table 2, the CRVM yields a sizable drop in the number of basis functions used as compared to CSVM. CRVM used only 70 and 121 relevance vectors for two different types of training-testing strategies while CSVM incorporates 523 and 594 support vectors for corresponding cases, respectively. This reduction in model complexity suggests that RVM should be effective in avoiding overfitting. Table 2 demonstrates substantial computation advantages of CRVM as compared with CSVM, where a significant reduction in learning time which carries the computation burden of method from 22 hours for CSVM to only 2.5 minutes for CRVM. These numbers are measured by evaluating algorithms on a Dell PC with Core Duo CPU at 2.66 GHz and 4 GB RAM.

Table 2. Comparison of CRVM to CSVM

Description	CRVM	CSVM
Number of design parameters	1	3
Number of design parameters combination	50	25000
Computing time	2.5 minutes	22 hours
Number of Basis Functions for training on 1 iteration	70	523
Number of Basis Functions for training on 5 iterations	121	594

4. CONCLUSION

We have shown potential advantages of a Channelized Relevance Vector Machine (CRVM) as a numerical observer for image quality assessment in the case of a myocardial perfusion defect detection task. The proposed method showed good ability to predict human observer scores accurately on test data different from the training set. Results have shown that CRVM has equal aggregate accuracy (AUC), if not better, than the previously tested CSVM method, while it drastically outperforms the CSVM in terms of generalization error (MSE). In addition, the CRVM has about 5 times lower model complexity (using 20% of the training samples as relevance vectors where CSVM used 99%). Also, CRVM has significantly lower computational cost corresponding to a smaller number of design parameters should be optimized via K-fold cross validation. Therefore, we have shown that the proposed CRVM method outperforms our previously introduced CSVM method with lower machine complexity, computational cost and number of basis functions.

ACKNOWLEDGEMENTS

This work was supported by National Institutes of Health under Grant HL65425 and Grant HL091017.

REFERENCES

[1] Myers, K. J. and Barrett, H. H., "Addition of a channel mechanism to the ideal-observer model," Journal of the Optical Society of America A 4, 2447–2457 (1987).

[2] Yao, J. and Barrett, H. H., "Predicting human performance by a channelized Hotelling observer model," in [Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series], Wilson, D. C. and Wilson, J. N., eds., Presented at the Society of Photo-Optical Instrumentation Engineers (SPIE) Conference 1768, 161–168 (1992).

[3] Abbey, C. K. and Barrett, H. H., "Human- and model-observer performance in ramp-spectrum noise: effects of regularization and object variability," Journal of the Optical Society of America A 18, 473–488 (2001).

[4] Brankov, J. G., Yang, Y., Wei, L., Naqa, I. E., and Wernick, M. N., "Learning a channelized observer for image quality assessment," IEEE Transactions on Medical Imaging 28, 991–999 (2009).

[5] Vapnik, V. N., [Statistical learning theory], Wiley, New-York, NY (1998).

[6] Tipping, M. E., "Sparse Bayesian learning and the relevance vector machine," Journal of Machine Learning Research 1, 211–244 (2001).

[7] Fletcher, T., "Relevance Vector Machines Explained," a tutorial on RVM (2008).

[8] Hanley, J. A., McNeil, B. J., "The meaning and use of the area under a receiver operating characteristic (ROC) curve," Radiology 143, 29-36 (1982).

[9] Barrett, H. H. and Myers, K. J., [Foundations of image science], Wiley-Interscience, Hoboken, NJ (2004).

[10] MOSEK ApS, "The MOSEK optimization software," [Available: http://www.mosek.com] (1997 [Accessed Apr 2010]).

[11] Weston, J., Elisseeff, A., Bakir, G., and Sinz, F., "The spider machine learning toolbox," [Available: http://www.kyb.tuebingen.mpg.de/bs/people/spider] (2005 [Accessed Apr 2010]).

[12] Narayanan, M. V., Gifford, H. C., King, M. A., Pretorius, P. H., Farncombe, T. H., Bruyant, P. P. and Wernick, M. N., "Optimization of iterative reconstructions of ^{99m}Tc cardiac SPECT studies using numerical observers," IEEE Transactions on Nuclear Science 49, 2355–2360 (2002).

[13] Pretorius, P. H., King, M. A., Tsui, B. M. W., LaCroix, K. J., and Xia, W., "A mathematical model of motion of the heart for use in generating source and attenuation maps for simulating emission imaging," Medical Physics 26, 2323–2332 (1999).

[14] Ljungberg, M. and Strand, S.-E., "A Monte Carlo program for the simulation of scintillation camera characteristics," Computer Methods and Programs in Biomedicine 29, 257–272 (1989).

[15] Hudson, H. M., and Larkin, R. S., "Accelerated image reconstruction using ordered subsets of projection data," IEEE Transaction on Medical Imaging 13.4, 601–609 (1994).