SUFFICIENT CONDITION FOR AN ADAPTIVE SYSTEM TO APPROXIMATE THE NEYMAN-PEARSON DETECTOR

Pilar Jarabo-Amores, Manuel Rosa-Zurera, Roberto Gil-Pita, Francisco López-Ferreras

Departamento de Teoría de la Señal y Comunicaciones,
Escuela Politécnica Superior, Universidad de Alcalá
Ctra. Madrid-Barcelona, km. 33.600, 28805, Alcalá de Henares - Madrid (SPAIN)
email: {mpilar.jarabo;manuel.rosa;roberto.gil;francisco.lopez}@uah.es

ABSTRACT

The application of adaptive systems to approximate the Neyman-Pearson detector is considered. The training error function is proved to be the key parameter that determines the possibility of approximating this detector. Based on the calculus of the approximated function for the selected error criterion, a sufficient condition is derived. Decision rules based on expressions of the optimum Bayes discriminant function, such as those approximated for the LMSE or the cross-entropy error criteria, have been analyzed. Previous works were based on the assumption that the system was trained to minimize the probability of error over the training set, so its performance was only optimal for the minimum probability of error threshold (system "operating point"). In this work, we prove that the decision rule based on the function approximated for an error function that fulfil the derived sufficient condition is optimum for all possible P_{FA} values. So, the concept of "operating point" will have no sense.

1. INTRODUCTION

The problem of approximating the Neyman-Pearson detector using an adaptive system trained in a supervised manner to minimize an error function, is considered. A theoretical study is presented that can be applied to any adaptive system, although many previous works focusses on neural networks.

Neural networks trained using the least mean squared-error (LMSE) and cross-entropy error criteria, have been applied to classification problems successfully.

In problems with more than two classes, a neural network with one output per class is usually used. If the desired value for the output associated to the class the input vector belongs to is 1, and 0 otherwise, the outputs of the neural network approximate the posterior probabilities of the classes [1, 2]. When using the cross-entropy error, the training strategy must guarantee that network outputs sum up to unity, while no care must be taken about it when using the LMSE criterion [3]. A minimum probability of error classifier can be approximated if the system decides in favor of the class associated to the highest output.

If there are only two classes, a neural network with only one output can be used. For desired outputs 1 for input vectors from class C_1, and -1 for input vectors from class C_2, Ruck et al. [1] demonstrated that the output of a neural network trained using the LMSE criterion approximates the function (1), which is an expression of the Bayes optimum discriminant function.

$$g_0(\mathbf{z}) = P(C_1|\mathbf{z}) - P(C_2|\mathbf{z}) \tag{1}$$

For desired outputs 1 for class C_1 and 0 for class C_2, the output of a network trained using the LMSE and cross-entropy error criteria approximates another expression of the Bayes optimum discriminant function:

$$g_1(\mathbf{z}) = P(C_1|\mathbf{z}) \tag{2}$$

V. Ramamurti, P.P. Gandhi et al. [4, 5] and J.L. Sanz and D. Andina [6, 7] applied multilayer perceptrons with one output to approximate the Neyman-Pearson detector. Once the neural network had been trained, they proposed to vary the detection threshold attending to P_{FA} requirements. In [4, 5] the LMSE criterion was used, while in [6, 7] the LMSE and cross-entropy errors were used, among others. For desired outputs 1 for hypothesis H_1 and 0 for hypothesis H_0, their works are based on the assumption that the network converges to an "operating point" that minimizes the probability of error over the training set. This point is not altered when adding a threshold detector to the network output with a threshold equal to 0.5, but when the detection threshold is varied attending to P_{FA} requirements, they argue the network will operate suboptimally.

In [8] a two outputs neural network was used, comparing the substraction of both outputs with a threshold, ap-

This work has been supported by the "Consejería de Educación de la Comunidad de Madrid" (SPAIN), under Project 07T/0036/2003 1

proach that is equivalent to that proposed in [1], although in this case the objective is to approximate the Neyman-Pearson detector.

These works were based on the assumption of an optimal "operating point" to combine the training without thresholding and the use of a threshold detector to approximate the Neyman-Pearson detector sub-optimally. So far, no attempt has been made to study the capability of neural networks, or any other type of adaptive system trained in a supervised manner, to approximate the Neyman-Pearson detector. In this paper, a theoretical study is presented to determine the conditions that must be satisfied.

2. PROBLEM FORMULATION

Perhaps, the cited works that used neural networks to approximate the Neyman-Pearson detector were conditional on the previous results obtained when applying such systems to classification problems. These results proved that neural networks could implement good approximations of the Bayes optimum discriminant function. Following a parallel reasoning, in this paper, a study of different expressions of the Bayes optimum discriminant function is carried out in order to determine their relationship with the Neyman-Pearson detector.

The Neyman-Pearson detector decision rule consists on comparing the likelihood ratio, or any equivalent discriminant function, to a detection threshold fixed attending to P_{FA} requirements. For a detection threshold equal to the ratio of prior probabilities, the decision rule based on the likelihood ratio is a realization of the minimum probability of error detector:

$$\Lambda(\mathbf{z}) = \frac{f(\mathbf{z}|H_1)}{f(\mathbf{z}|H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \frac{P(H_0)}{P(H_1)} \tag{3}$$

After simple manipulations, rule (3) can be expressed as:

$$g_0(\mathbf{z}) = P(H_1|\mathbf{z}) - P(H_0|\mathbf{z}) \underset{H_0}{\overset{H_1}{\gtrless}} 0 \tag{4}$$

$$g_1(\mathbf{z}) = P(H_1|\mathbf{z}) \underset{H_0}{\overset{H_1}{\gtrless}} 0.5 \tag{5}$$

$$g_2(\mathbf{z}) = P(H_1)f(\mathbf{z}|H_1) - P(H_0)f(\mathbf{z}|H_0) \underset{H_0}{\overset{H_1}{\gtrless}} 0 \tag{6}$$

The equivalence of expressions (3), (4), (5) and (6), can be proved taking into consideration the following relations:

$$P(H_1|\mathbf{z}) = \frac{f(\mathbf{z}|H_1)P(H_1)}{f(\mathbf{z})} \tag{7}$$

$$f(\mathbf{z}) = f(\mathbf{z}|H_1)P(H_1) + f(\mathbf{z}|H_0)P(H_0) \tag{8}$$

Note that rule (4) is based on the discriminant function approximated by neural networks trained using the LMSE criterion with desired outputs $\{-1, 1\}$, and rule (5) is based on the one approximated with desired outputs $\{0, 1\}$, using the LMSE and the cross-entropy error criteria. The question that arises is: for any other threshold value, can these discriminant functions be used to implement the Neyman-Pearson detector? In next section, we answer this question and determine a sufficient condition that a discriminant function must fulfill in order to be used to implement the Neyman-Pearson detector for any pair of likelihood functions.

3. PROPOSED METHOD

For answering the question formulated in the previous section, the following method is proposed:

1. Modify expressions (4),(5) and (6) replacing the fixed thresholds with variable ones. These new thresholds will be denoted as η_0, and will be determined attending to P_{FA} requirements.

2. Extract the likelihood ratio from the decision rules obtained in the previous step, identifying the detection thresholds required for the decision rules based on the likelihood ratio, η_{lr}.

3. Check if η_{lr} does not depend on the input vector.

When this method is applied to rules (4), (5) and (6), the following results are obtained:

$$\Lambda(\mathbf{z}) \underset{H_0}{\overset{H_1}{\gtrless}} \eta_{lr} = \frac{P(H_0)(1+\eta_0)}{P(H_1)(1-\eta_0)} \tag{9}$$

$$\Lambda(\mathbf{z}) \underset{H_0}{\overset{H_1}{\gtrless}} \eta_{lr} = \frac{\eta_0 P(H_0)}{P(H_1)(1-\eta_0)} \tag{10}$$

$$\Lambda(\mathbf{z}) \underset{H_0}{\overset{H_1}{\gtrless}} \eta_{lr} = \frac{\eta_0 f^{-1}(\mathbf{z}|H_0) + P(H_0)}{P(H_1)} \tag{11}$$

From rules (9), (10) and (11), expressions of the required detection thresholds, η_0, can be derived, once the likelihood ratio thresholds, η_{lr}, have been fixed attending to P_{FA} conditions.

Rules (9) and (10) reveal that the discriminant functions $g_0(\mathbf{z})$ and $g_1(\mathbf{z})$ can be used to implement the Neyman-Pearson detector for any pair of likelihood functions, because for each value of P_{FA}, a detection threshold, η_0 can be determined that is only a function of problem parameters.

Rule (11) is also obtained from a decision rule based on an expresion of the optimum Bayes discriminant function, but there is no guarantee that a detection threshold η_0 independent of input vector can be found for any pair of likelihood functions.

This method allows us to prove that a neural network, or any other adaptive system trained to approximate the discriminant functions $g_0(\mathbf{z})$ or $g_1(\mathbf{z})$, can be used to approximate the Neyman-Pearson detector, adding a threshold detector to the system output, and varying the threshold attending to P_{FA} requirements. When varying the threshold, the system performance will not be suboptimal, because the approximated discriminant function can be used to implement the optimum detector. For now on, we have proved that the concept of neural network "operating point" have no sense.

In next section, the decision rule based on $g_2(\mathbf{z})$ is studied, proving that the proposed method defines a sufficient, but not necessary, condition for a discriminant function being suitable for implementing the Neyman-Pearson detector.

4. EXAMPLE

As an example, the discriminant function $g_2(\mathbf{z})$ defined in (6) is applied to detect gaussian signals in white gaussian noise. We have proved in section 3, that there is no guarantee that it can be used to implement the Neyman-Pearson detector for any pair of likelihood functions.

Two cases of study are proposed, one for which $g_2(\mathbf{z})$ can be used to implement such a detector, and another one for which this is not a suitable function.

As white gaussian noise is assumed, under H_0, $\mathbf{z}$ is a vector of N zero mean and variance σ_n^2 independent gaussian random variables. Under H_1 they have zero mean, variance $\sigma_n^2 + \sigma_s^2$ and covariance matrix $\mathbf{C}$. The signal-to-noise ratio is defined as:

$$snr = \sigma_s^2 / \sigma_n^2 \tag{12}$$

Assuming that $\sigma_n^2 = 1$, the signal-to-noise ratio is equal to σ_s^2, that is expressed in decibels as $SNR = 10\log_{10}(snr)$.

The likelihood functions under each hypothesis are:

$$f(\mathbf{z}/H_0) = \frac{1}{\sqrt{(2\pi)^N}} \exp(-\frac{1}{2}\mathbf{z}^T\mathbf{z}) \tag{13}$$

$$f(\mathbf{z}/H_1) = \frac{1}{\sqrt{(2\pi)^N \det(\mathbf{C})}} \exp(-\frac{1}{2}\mathbf{z}^T\mathbf{C}^{-1}\mathbf{z}) \tag{14}$$

where $\det(\mathbf{C})$ gives the determinant of $\mathbf{C}$.

The snr selected for designing the detector (the snr value used to calculate $\mathbf{C}$ in (14)), will be denoted as snr_d, or SNR_d in decibels. On the contrary, the snr of the input vector, $\mathbf{z}$, will be denoted as snr.

For $N = 16$ and $P(H_0) = P(H_1) = 0.5$, two cases are considered.

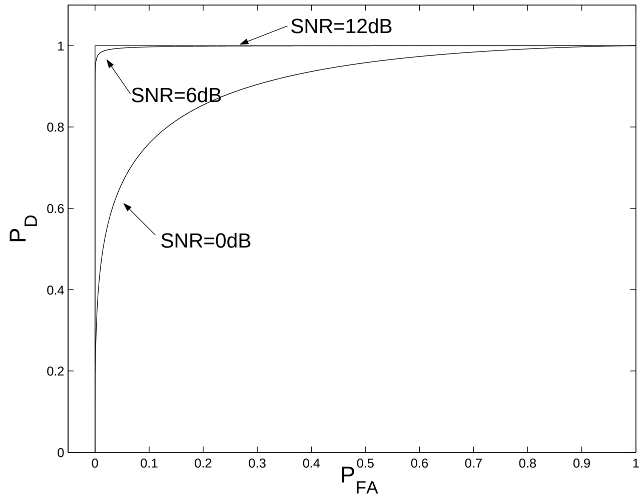

Fig. 1. ROC curves for the detection rule based on $g_2(\mathbf{z})$ with $\mathbf{C} = (snr_d + 1)\mathbf{I}$, $SNR_d = 6dB$ and different SNRs (solid line) and Neyman-Pearson detector ones (dashed lines).

4.1. Case I: C is a scalar multiple of the identity matrix

For $\mathbf{C} = (snr_d + 1)\mathbf{I}$, where $\mathbf{I}$ is de $(N\text{x}N)$ identity matrix, the decision rule based on $g_2(\mathbf{z})$ can be expressed as:

$$\frac{1}{2\sqrt{(2\pi)^N(snr_d+1)^N}} \exp(-\frac{\mathbf{z}^T\mathbf{z}}{2(snr_d+1)}) - \frac{1}{2\sqrt{(2\pi)^N}} \exp(-\frac{1}{2}\mathbf{z}^T\mathbf{z}) \underset{H_0}{\overset{H_1}{\gtrless}} \eta_0 \tag{15}$$

Fixing η_0 attending to P_{FA} requirements, we have calculated the P_D for $SNR_d = 6dB$ and different SNR values. ROC curves for $SNR = 0, 6, 12$ dB are presented in figure 1, which are superimposed on the Neyman-Pearson detector ones (in dashed line), showing that this rule is an implementation of the Neyman-Pearson detector.

Applying simple manipulations on (15), a simpler equivalent discriminant function can be found, consisting of the squared norm of the input vector. Taking into consideration that this statistic is independent of snr_d, that the P_{FA} is evaluated under hypothesis H_0 and that the noise variance remains constant, we can conclude that this detector is snr_d independent.

The obtained results can be explained dividing both terms of (15) by $f(\mathbf{z}|H_0)$, an obtaining an equivalent decision rule based on the likelihood ratio:

$$\frac{1}{2}\Lambda(\mathbf{z}) - \frac{1}{2} \underset{H_0}{\overset{H_1}{\gtrless}} \eta_0 f^{-1}(\mathbf{z}|H_0) \tag{16}$$

Taking into consideration (17), expression (16) can be rewritten as in (18). Both terms of rule (18) are constant on hyper-spheres centered on the origin. Because of that, a detection threshold η_0' that does not depend on the input vector can be determined for each P_{FA}.

$$\Lambda(\mathbf{z}) = \frac{1}{(snr_d+1)^{N/2}} \exp\left(\frac{snr_d}{2(snr_d+1)}\mathbf{z}\mathbf{z}^T\right) \quad (17)$$

$$\frac{1}{2(snr_d+1)^{N/2}} \exp\left(\frac{snr_d}{2(snr_d+1)}\mathbf{z}\mathbf{z}^T\right) - \frac{1}{2} \underset{H_0}{\overset{H_1}{\gtrless}} \eta_0 (2\pi)^{N/2} \exp\left(\frac{\mathbf{z}\mathbf{z}^T}{2}\right) = \eta_0' \quad (18)$$

4.2. Case II: C is not a scalar multiple of the identity matrix

In this case, the covariance matrix under hypothesis H_1 is given by:

$$\mathbf{C} = \begin{pmatrix} \mathbf{I} + snr_d\mathbf{U} & \mathbf{O} \\ \mathbf{O} & \mathbf{I} + snr\mathbf{U} \end{pmatrix} \quad (19)$$

$\mathbf{U}$ is a $(\frac{N}{2}\text{x}\frac{N}{2})$ unit matrix and $\mathbf{O}$ is a $(\frac{N}{2}\text{x}\frac{N}{2})$ matrix of zeros.

ROC curves for $SNR_d = 6dB$ and $SNR = 0, 6, 12$ dB are presented in figure 2, where we can observe that the detector based on $g_2(\mathbf{z})$ curve and the Neyman-Pearson one are tangent in one point.

For $SNR = SNR_d$, the decision rules (3) and (6) are equivalent, the tangential point corresponds to the minimum probability of error. For other values of SNR, as P_{FA} is evaluated under hypothesis H_0, and noise variance remains constant, the P_{FA} is the same, but the P_D will be a function of SNR.

Taking into consideration that for each SNR_d value, the discriminant function of rule (6) is different, the pair of P_{FA} and P_D values associated to the minimum probability of error are different. Because of that, the tangential points for $SNR \neq SNR_d$ does not correspond to the minimum probability of error for the selected SNR. This reasoning allows us to conclude that the minimum probability of error detector is SNR_d dependent.

Attending to the sufficient condition derived in this paper, a study is carried out about the possibility of finding a equivalent decision rule based on the likelihood ratio, with a detection threshold independent on the input vector. The Neyman-Pearson detector decision rule based on the likelihood ratio is given in (20), where $h(\mathbf{z})$ is defined in (21).

$$\Lambda(\mathbf{z}) = \frac{1}{1+\frac{N}{2}snr_d} \exp\left[\frac{snr_d}{2(\frac{N}{2}snr_d+1)}h(\mathbf{z})\right] \underset{H_0}{\overset{H_1}{\gtrless}} \eta_{lr} \quad (20)$$

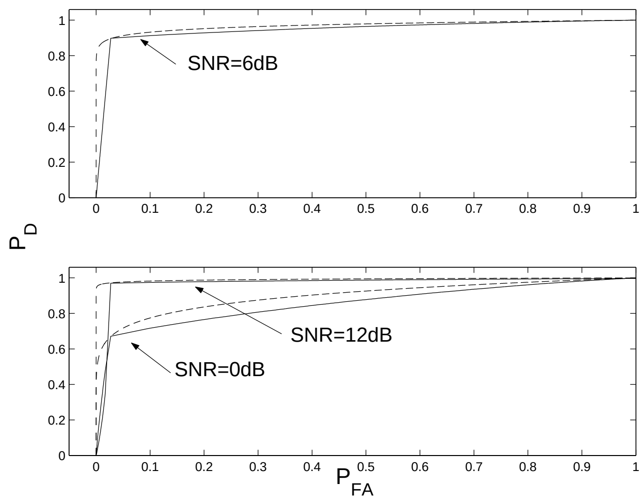

Fig. 2. ROC curves for the detection rule based on $g_2(\mathbf{z})$ with $\mathbf{C}$ defined in (19), $SNR_d = 6dB$ and different SNRs (solid line) and Neyman-Pearson detector ones (dashed lines).

$$h(\mathbf{z}) = \left(\sum_{i=1}^{N/2} z_i\right)^2 + \left(\sum_{i=1+N/2}^{N} z_i\right)^2 \quad (21)$$

Making simple mathematical manipulations, rule (20) can be proved to be equivalent to (22), which reveals that $h(\mathbf{z})$ is a sufficient statistic. This statistic is independent of snr_d. As the noise variance remains constant, and the detection threshold is fixed attending to P_{FA} requirements, this detector is snr_d independent, in contrast to the minimum probability of error one.

$$h(\mathbf{z}) \underset{H_0}{\overset{H_1}{\gtrless}} \frac{2(\frac{N}{2}snr_d+1}{snr_d} ln[\eta_{lr}(1+\frac{N}{2}snr_d)] = \eta_h \quad (22)$$

Substituting the likelihood ratio in (16), the decision rule (23) is obtained. In this case, the surfaces where both terms of the decision rule are constant, are quite different. So, a η_0' value independent of $\mathbf{z}$ can not be found for any P_{FA}.

$$\frac{1}{2(1+\frac{N}{2}snr_d)} \exp\left[\frac{snr_d}{2(\frac{N}{2}snr_d+1)}h(\mathbf{z})\right] - \frac{1}{2} \underset{H_0}{\overset{H_1}{\gtrless}} \eta_0(2\pi)^{N/2}\exp\left(\frac{\mathbf{z}\mathbf{z}^T}{2}\right) = \eta_0' \quad (23)$$

5. CONCLUSIONS

In this paper, the possibility of approximating the Neyman-Pearson detector using adaptive systems trained in a supervised manner is studied. Previous works that used neural networks, were based on the hypothesis that the network was trained to minimize the probability of error over the training set. So, for a detection threshold different from the minimum probability of error one, the network was supposed to be forced to work sub-optimally.

Taking the discriminant functions approximated using the LMSE and the cross-entropy error criteria as a starting point, decision rules based on different expressions of the optimum Bayes discriminant function have been analyzed. A sufficient condition has been determined to guarantee a discriminant function can be used to implement the Neyman-Pearson detector for any pair of likelihood functions. We have proved that if the likelihood ratio can be extracted from the decision rule based on the discriminant function approximated by the adaptive system, obtaining an equivalent one with a detection threshold independent of the input vector, the system output can be used to approximate the Neyman-Pearson detector for any P_{FA} value. So, the key parameter is the error function selected for training, because it determines the approximated function.

When the presented results are applied to previous works related with neural networks, if the approximated function associated to the selected error function fulfills the derived condition, the system performance will not have to be suboptimal for P_{FA} values different from the minimum probability of error one, and the concept of "operating point" will have no sense.

6. REFERENCES

[1] D.W. Ruck, S.K. Rogers, M. Kabrisky, M.E. Oxley, and B.W. Suter, "The multilayer perceptron as an aproximation to a bayes optimal discriminant function," *IEEE Transactions on Neural Networks*, vol. 1, no. 4, pp. 296–298, December 1990.

[2] E.A. Wan, "Neural network classification: a bayesian interpretation," *IEEE Transactions on Neural Networks*, vol. 1, no. 4, pp. 303–305, December 1990.

[3] C. M. Bishop, *Neural Networks for Pattern Recognition*, Oxford University Press, U.K., 1 edition, 1995.

[4] V. Ramamurti, S. S. Rao, and P. P. Gandhi, "Neural detectors for signals in non-gaussian noise," in *Proceedings of the IEEE International Conference Acoustics, Speech, and Signal Processing*, Minneapolis, MN USA, April 1993, vol. 1, pp. 481–484.

[5] P.P. Gandhi and V. Ramamurti, "Neural networks for signal detection in non-gaussian noise," *IEEE Transactions on signal processing*, vol. 45, no. 11, pp. 2846–2851, November 1997.

[6] D. Andina, J.L. Sanz-Gonzlez, and J.A. Jimnez-Pajares, "A comparison of criterion functions for a neural network applied to binary detection," *IEEE International Conference on Neural Networks*, vol. 1, pp. 492–495, December 1995.

[7] D. Andina, J.L. Sanz-Gonzlez, and O.A. Rodrguez-Martn, "Performance improvements for a neural network detector," *IEEE International Conference on Neural Networks*, vol. 1, pp. 492–495, December 1995.

[8] D.J. Munro, O.K. Ersoy, M.R. Bell, and J.S. Sadowsky, "Neural network learning of low probability events," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 32, no. 3, pp. 898–910, 1996.