NEURAL NET CLASSIFICATION OF THE γ- AND P- IMAGES REGISTERED WITH ATMOSPHERIC CHERENKOV TECHNIQUE, RANDOM SEARCH LEARNING IN FEED-FORWARD NETWORKS

OG 4.7.23

A.A.CHILINGARIAN

Yerevan Physics Institute, Yerevan, Republic Armenia

Abstract

A new method of data analysis, based on mathematical models of Neural Net, is applied for background rejection in γ-ray atronomy experiments. The results illustrated the advantages of the proposed technique compared to Bayesian approach.

Introduction. The term "learning" is applied here to neural net mathematical models (artificial neural nets), mimicking the most complicated and most important types of biological activity - classification and pattern recognition.

Learning is a key in approaching the excellent possibilities of biological nets and if we can organize learning procedure, comparable to that one in biological nets, we can achieve the goal - to construct a fast intelligent classifier.

The random search is an universal powerful methodology, akin the trial and error method, and may be it is the basis of the unpredictable efficiency of biological neural nets.

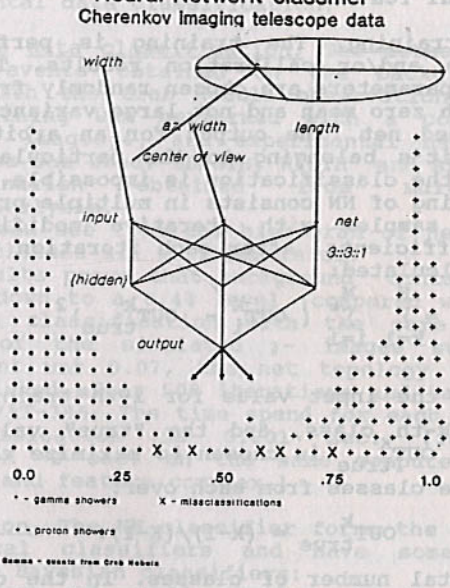
We use random search algorithms to train the neural net for discrimination studies in very high energy astrophysics. The problem chosen - the reduction of the the background hadron contamination in detecting the γ-rays from point sources is one of the most important problems of γ-ray astronomy.

In our previous works, learning was performed in the framework of the Bayesian paradigm, by multidimensional a posteriori probability density estimation. This method was strongly dependent on the choice of the particular nonparametric method of the density estimation with it's free parameters, and was rather time consuming.

The Neural Nets were trained on the same samples and so, for the first time, we compare the two alternative classification techniques on experimental data, thus providing the continuity in development of new information technologies.

The Neural Net models. The Neural Net(NN) is defined by specifying the total number of neurons (nodes) and layers, and the connections among the neurons. The input layer of the network have one node for each feature, signal processing is performed layer by layer beginning from the input (see fig. 1). Neurons of successive layers receive input only from neurons of the previous layer and each neuron in a given layer sends it's output to all

nodes in the next layer. The output layer have a single node, determining the value of the discrimination function.



The input of each neuron (excepts the first input layer) consists of the weighted sum of outputs of previous layer neurons:

$$IN_i^{l+1} = \sum_{j=1}^{NODES(l)} J_{ij} OUT_j^l , \quad j = 1, NODES(l+1), \quad l = 1, L, \quad (1)$$

where l is layer index, L- the total number of layers, NODES(l) - the neurons number in l-th layer, OUT_j^l- the output of j-th neuron in l-th layer. J_{ij}- the connection strength (weight) of i-th to j-th neurons.

The output of the neuron is formed by a nonlinear, so called, sigmoid function:

$$OUT_i^l = 1/(1-\exp(-IN_i^l)), \quad i = 1, NODES(l), \quad l \neq 1, \quad (2)$$

where IN_i^l is the input of i-th neuron in l-th layer.

With an input/output relation thus defined, the multidimensional feature set is translated from input through hidden layers to the output node, where classification is performed. So, the NN provides one to one mapping of complicated input signal to class assignments.

Such a data handling design, combining the linear

summation on the nodes input, and nonlinear transformation in the nodes, allows us to take into account all distinctive information, including differences in nonlinear correlations of alternative classes of multidimensional features.

The NN training. The training is performed with simulated data and/or calibration results. The initial values of net parameters are chosen randomly from Gaussian population with zero mean and not large variance. In such, randomly trained net, the output for an arbitrary input set, despite it's belonging to the particular type, is near 0.5, and the classification is impossible.

The training of NN consists in multiple processing of all training samples with iterative modifications of connection coefficients. After each iteration the quality function is calculated:

$$Q = \sum_{k=1}^{K} \sum_{i=1}^{M_k} (OUT_i^k - OUT_{true}^k)^2 , \qquad (3)$$

where OUT_i^k is the input value for i-th training sample, belonging to k-th class, and the "true" value of k-th class output - OUT_{true}^k is chosen to maximize the shift of the alternative classes from each over:

$$OUT_{true}^k = (k-1)/(K-1). \qquad (4)$$

where K is total number of classes. In the case of two classes, the "true" outputs, as one can easily see, are equal to zero and one.

The quality function minimization usually is done by the back propagation method[2] - gradient descend is performed on the quality function with respect to the weights in order to minimize the deviations of the network response from the desired "true" response. The main obstacle of such methods is their convergence to a local minimum, in contrast to the random search in net parameters space which allows one to escape from local minimum region and continue the search till global minimum is reached.

The total number of searching net parameters equals:

$$NTOT = \sum_{l=2}^{L} NODES(l) + \sum_{l=1}^{L-1} NODES(l)\, NODES(l-1). \qquad (5)$$

For simple net configurations 1::3::1 - NTOT= 10, for - 3::3::1 - 18.

First the particular net parameter is randomly chosen, then the random addition (or subtraction) Δ is randomly selected:

$$\Delta \approx \text{const}\ f(Q)\ (RNDM - 0.5), \qquad (6)$$

where RNDM - is randomly distributed in the (0 - 1) interval, f(Q) is the power function controlling the

random step size during the process of reaching the global minimum and CONST is a normalizing factor.

The iterations cancel when the value of the quality function is stabilized, and no more improvements take place. Then the obtained set of net parameters can be used for experimental data classification.

The Crab data classification. The method was tested using 5000 events obtained on the background region associated with the Crab Nebula observation of ref.[3].

The training was performed with a combination of simulated γ- images[4] and experimental hadron images. Three features was used: LENGTH, WIDTH and AZWIDTH - the best combination obtained from multidimensional correlation analysis.

As one can see from the histogram of net output the alternative classes all well separated.

The results prove that background contamination can be rejected down to a 0.4% level (compared with 0.6% for the Bayesian classification with the same sample[5]). Nearly 50% of the simulated γ- images survived. The decision point was 0.07, the net topology was 3::3::1. Training finished after 500 iterations and takes ≈10 min. on an IBM PC/AT-386. The time spend for each experimental event classification is 0.001 sec. (for Bayesian classification ≈2 sec. on the same computer with same training set and feature complex.)

Conclusion. The NN classifier forms the special type of statistical classifiers and have some advantages comparing the Bayesian classifiers:

the effect of small samples, the main obstacle deteriorated all nonparametric methods, is not of crucial importance for NN,

NN classifiers use the training sample more effectively (all samples for each coefficient); in the Bayesian approach each class was trained only with own training sample.

NN classifiers are very fast - 3 order of magnitude faster than Bayesian classifiers.

We can therefore recommend NN classifiers not only for off-line analysis, but for on-line analysis in collider and cosmic ray physic experiments as very fast intelligent trigger. For the air Cherenkov technique, the NN trigger can be used instead of an arbitrary hardware trigger, providing optimal event selection and the incorporating of the events with large impact parameters.

References

1. Chilingarian A.A., Comp. Phys Comm. 54, 381-390, (1989).
2 Hecht-Nielsen, Proc. Int. Joint. Conf. on Neural Networks, v.1, pp. 593-605, Washington,(1989).
3. Vacanti G. et al, Ap. J 377, 467, (1991).
4.Macomb D.J. and Lamb R.C., Proc. 21 Int. Cosmic Ray Conf. 2, 445, Adelaide, (1990).
5. Cawley M.F., Chilingarian A.A., Proc. 22 Int. Cosmic Ray Conf., Dublin, OG-4.7.2, (1991).