ANN AS SUBJECT OF STATISTICAL TESTING

J. Kukal[1], *A. Bartoš*[2,3], *D. Řípová*[3], *D. Majerová*[4]

[1]ICT Prague, [2]UHKV Prague, [3]PCP Prague, [4]CTU Prague

Abstract

The paper is oriented to statistical testing of ANN output in special cases. The neural network with several inputs and single output is able to classify objects into two classes depending on the ANN output value. Basic types of neural networks are included: linear perceptron, bipolar perceptron, sigmoidal perceptron, MLP and RBF. Various statistical properties of ANN output signal enforce various strategies of statistical testing. The non-parametric approach is represented by Fisher's factorial, chi square, Wilcoxon-Mann-Whitney and Kolmogorov-Smirnov test. The general methodology is based on ANN learning on the training pattern set and statistical testing on the verification set. All the calculations were performed in the Matlab environment.

1 Introduction

Artificial Neural Network is a modern tool for data processing and categorization. When the ANN was learned on training pattern set, it could be tested on another verification pattern set. The difference between ANN output and its given value can be subject of statistical testing. The hypothesis testing depends on the character of ANN output. When the ANN input is a stochastic vector, the ANN output is also stochastic variable with known or unknown distribution. The are five types of single output neural networks: linear perceptron, bipolar perceptron, sigmoid perceptron, MLP, RBF and their output can be subject of statistical testing. It is necessary to introduce the basic ANN types first.

2 ANN Preliminaries

Let $n \in \mathbf{N}$, input vector $\mathbf{x} = (x_1, ..., x_n) \in \mathbf{R}^n$ and output value $y^* \in \mathbf{D} \subseteq \mathbf{R}$. Then the vector $\mathbf{p} = (\mathbf{x}, y^*)$ is called *pattern* in this context. The traditional output domains are $\mathbf{D}_1 = \{0;1\}$, $\mathbf{D}_2 = \{-1;+1\}$, $\mathbf{D}_3 = [0;1]$, $\mathbf{D}_4 = [-1;+1]$ or $\mathbf{D}_5 = \mathbf{R}$. Let $m \in \mathbf{N}$. Then $\mathbf{PS} = \{\mathbf{p}_1, ..., \mathbf{p}_m\}$ is called *pattern set*. In the special case of $\mathbf{D} = \mathbf{D}_2$, the pattern set can be split into subsets $\mathbf{PS}^+ = \{\mathbf{p} \in \mathbf{PS} \mid y^* = +1\}$ and $\mathbf{PS}^- = \{\mathbf{p} \in \mathbf{PS} \mid y^* = -1\}$, which are called sets of positive and negative patterns. The pattern set can be also split into training and verification pattern sets and their subparts. The *training pattern set* and its positive and negative parts are denoted as $\mathbf{TS}, \mathbf{TS}^+, \mathbf{TS}^-$. The *verification pattern set* and its positive and negative parts are denoted as $\mathbf{VS}, \mathbf{VS}^+, \mathbf{VS}^-$.

The a*rtificial neural network* (ANN) with single output can be represented as a function $\mathrm{ANN} : \mathbf{R}^n \to \mathbf{R}$ and its output is then expressed as $y = \mathrm{ANN}(\mathbf{x})$. There are many various models of ANN with single output. The direct processing comes to two-layer systems as *linear*, *bipolar* and *sigmoid perceptrons*, which are represented by formulas

$$y_{\mathrm{LIN}} = w_0 + \sum_{k=1}^{n} w_k x_k$$

$$y_{\mathrm{BIP}} = \mathrm{sign}\left(w_0 + \sum_{k=1}^{n} w_k x_k \right)$$

$$y_{\mathrm{SIG}} = \tanh\left(w_0 + \sum_{k=1}^{n} w_k x_k \right)$$

with weights as subjects of learning on the training pattern sets. Introducing the hidden layer of $H \in \mathbf{N}$ artificial neurons, we obtain three-layer ANN of $n - H - 1$ topology. According to the literature the most frequent ANN's are *multilayer perceptron* (MLP), *MLP with linear output* (MLL) and *radial basis function network* (RBF). The adequate formulas with unknown weights are

$$y_{\mathrm{MLP}} = \tanh\left(v_0 + \sum_{k=1}^{H} v_k \tanh\left(w_{k,0} + \sum_{j=1}^{n} w_{k,j} x_j \right)\right)$$

$$y_{\mathrm{MLL}} = v_0 + \sum_{k=1}^{H} v_k \tanh\left(w_{k,0} + \sum_{j=1}^{n} w_{k,j} x_j \right)$$

$$y_{\mathrm{RBF}} = v_0 + \sum_{k=1}^{H} v_k \exp\left(- \frac{\sum_{j=1}^{N} \left(x_j - w_{k,j} \right)^2}{2 w_{k,0}^{\;2}} \right)$$

3 Statistical testing

Proposed methods of statistical testing can be applied only in a special case of verification set of bipolar patterns. Thus, the training pattern set cannot be used for the testing to prevent the incorrectness. It implies that the ANN weights can be estimated from the training pattern set using any learning method but they must be independent on the verification set of patterns.

3.1 Bipolar testing

Beginning with learned bipolar perceptron, we can form 2×2 contingency table with individual, marginal and total event frequencies. There are only four possible events and the contingency table consists of values

	y = -1	y = +1	
y^* = -1	a	b	P
y^* = +1	c	d	Q
	R	S	N

Supposing that $\min(b,c) \le \min(a,d)$, we can use *Fisher factorial test* and evaluate the testing criterion

$$p = \sum_{\Delta=0}^{\min(b,c)} \frac{P!Q!R!S!}{N!(a+\Delta)!(b-\Delta)!(c-\Delta)!(d+\Delta)!}$$

with critical value 0.05. Non-passing the test means the success in ANN learning.

In the case of significant marginal frequencies when $\frac{\min(P,Q)\min(R,S)}{N} \ge 5$, we can apply χ^2 *test* with testing criterion $\chi^2 = N \frac{(ad-bc)^2}{PQRS}$ which belongs to χ^2 distribution with one degree of freedom. The high value of testing criterion is an indicator of ANN learning quality.

The previous two tests can be also used for another ANN types after the output thresholding. Let $\theta \in \mathbf{R}$ be threshold value, which was obtained via training pattern set. The bipolar output variable can be easily obtained as $y = \mathrm{sign}(y_{\mathrm{ANN}} - \theta)$. It implies that linear perceptron can substitute bipolar or sigmoidal ones in this type of tests. In analogy, the MLP is equivalent to MLP with linear output in this case.

3.2 Rank based testing

Continuous output of ANN can be investigated via rank based testing. Let $m = \left|\mathbf{VS}^{+}\right|$, $n = \left|\mathbf{VS}^{-}\right|$ be sizes of two ANN *output samples* $(\xi_1,...,\xi_m)$ and $(\eta_1,...,\eta_n)$. Let $(\psi_{(1)},...,\psi_{(m+n)})$ be *ordered union sample* and r_k be *rank* of ξ_k in it.

Let $W = \sum_{k=1}^{m} r_k$. The *Wilcoxon-Mann-Whitney test* is based on testing criterion

$$U = mn + \frac{m(m+1)}{2} - W \ \text{ with } \mathrm{E}U = \frac{mn}{2}, \quad \mathrm{D}U = \sqrt{\frac{mn}{12}(m+n+1)}$$

Another approach is based on the distance between experimental cumulative distribution functions $\mathrm{F}_m(x), \mathrm{F}_n(x)$ of ANN output samples. The *Kolmogorov-Smirnov test* evaluates critical probability

$$p = 1 - \mathrm{F}_{\mathrm{KS}}\left(D\sqrt{\frac{mn}{m+n}} \right) \text{ with } D = \sup_{x \in \mathbf{R}} \left| \mathrm{F}_m(x) - \mathrm{F}_n(x) \right|$$

and cumulative distribution function $\mathrm{F}_{\mathrm{KS}}(x) = 1 - 2\sum_{k=1}^{\infty} (-1)^{k-1} \exp(-2k^2 x^2)$.

The sigmoid function is increasing one. It implies the testing comes to the same results for linear and sigmoid perceptron. The MLP is also equivalent to MLL output in the same meaning.

4 Illustrative example

A MLL neural network with 12 inputs 4 hidden neurons and one output were trained on 33 positive and 14 negative patterns. The statistical properties was studied in the case of 16 positive and 16 negative patterns with threshold value $\theta = 0$.

Fisher's factorial test with $a = 16, b = 0, c = 1, d = 15$ offered the actual probability $p = 2.828 \times 10^{-9} < 0.05$, which means significant dependence between ANN output and pattern output.

The second test of contingency table came to $\chi^2 = 28.235$ and thus $p = 1.074 \times 10^{-7} < 0.05$, which also means significant dependence.

The Wilcoxon-Mann-Whitney test evaluated the criterion $U = 16$, which implies $p = 2.431 \times 10^{-5} < 0.05$ and the symmetry of union sample is refused.

The value $D = 0.9375$ was obtained via Kolmogorov-Smirnov test. Thus $p = 3.327 \times 10^{-9} < 0.05$, which means the samples are not from the same distribution.

5 Conclusions

Various statistical approaches were used for the testing of ANN output on the verification set. The non-parametric techniques based on contingency tables, ranks, symmetry or cumulative distribution function distance, play the role in correct statistical testing of ANN output. Given example exhibits the high sensitivity of Fisher's factorial test and Kolmogorov-Smirnov test. But the Wilcoxon-Mann-Whitney test was not too sensitive in the case of strongly separated samples. Proposed methodology enables an alternative view to the quality of ANN learning.

Acknowledgements

The work has been supported by the Ministry of Education, Youth and Sport of the Czech Republic No. 1M0002375201. This support is very gratefully acknowledged.

References

[1] Press W.H., Tuekolsky S.A., Vetterling W.T., Flannery B.P.: *Numerical Recipes in C*, Cambridge University Press, Cambridge, 1992.

[2] Riley K.F., Hobson M.P., Bence S.J.: *Mathematical Methods for Physics and Engineering*, Cambridge University Press, Cambridge, 1992.

[3] Meloun M., Militký J.: *Statistické zpracování experimentálních dat*, PLUS, Praha, 1994.

[4] Olehla M., Věchet V., Olehla J.: *Řešení úloh matematické statistiky ve Fortranu*, NADAS, Praha, 1982.

[5] Anděl J.: *Matematická statistika*, SNTL, Praha, 1978.

Jaromír Kukal

Department of Computing and Control Engineering, Institute of Chemical Technology
Technická 5, 166 28 Prague 6 – Dejvice
phone: +420 220 444 212, fax: +420 220 445 053, e-mail: jaromir.kukal@vscht.cz