Design of order statistics filters using feedforward neural networks

Yu.S. Maslennikova, V.V. Bochkarev

Kazan Federal University, Kremlevskaya str.18, Kazan 420018, Russia

E-mail: JSMaslennikova@kpfu.ru, vbochkarev@mail.ru

Abstract. In recent years significant progress have been made in the development of nonlinear data processing techniques. Such techniques are widely used in digital data filtering and image enhancement. Many of the most effective nonlinear filters based on order statistics. The widely used median filter is the best known order statistic filter. Generalized form of these filters could be presented based on Lloyd's statistics. Filters based on order statistics have excellent robustness properties in the presence of impulsive noise. In this paper, we present special approach for synthesis of order statistics filters using artificial neural networks. Optimal Lloyd's statistics are used for selecting of initial weights for the neural network. Adaptive properties of neural networks provide opportunities to optimize order statistics filters for data with asymmetric distribution function. Different examples demonstrate the properties and performance of presented approach.

1. Introduction

In recent times, many linear filtering techniques were widely used in various fields of science and technology. At the same time, the use of linear filtering does not provide a satisfactory solution for many practical applications. It is known, for example, that the problem of optimal filtering can be solved in the class of linear filters when the additive noise is statistically independent and has a distribution function that is close to the normal distribution. In practice, the additive noise can be dependent on desired signal or have a distribution function the is different from the normal. In these cases, the optimal solution can be found in the class of non-linear filters, for example, using artificial neural networks [1].

Linear order statistics (or linear order estimation of some variable μ) is a combination $\hat{\mu} = \sum_{i=1}^{n} \alpha_i x^{(i)}$, where $x^{(i)}$ - i-th term in an ordered (ranked) series. Particularly, the arithmetic average can be obtained as an example of the linear order statistic by setting coefficients $\alpha_i = 1/n$ [2]. Some cases of order statistics play a major role in the various statistical studies, such as extrema, sampling range, sample median and other.

The problem of optimal weighting of observations (or finding the optimal linear statistics) was set in the middle of the last century, and soon was successfully solved by E.H. Lloyd, who found a common solution (on the assumption that the distribution function is known) for the evaluation of the two unknown parameters of the distribution function: mean and variance [3]. However, in practice, such linear estimates are insufficient. Therefore, in recent years nonlinear order statistics filters have been developed, especially for digital image processing. Authors of the paper [4, 5] present filters that is based on nonlinear order statistics using perceptron.

In the present work non-linear analogues of order statistics based on feedforward neural networks is proposed. The main idea of the proposed technique is based on Lloyd's statistics, that are used for initialization of the neural network weights. Nonlinear activation function of neurons improves the quality of filtering results in cases with distribution different from the normal.

2. Lloyd's statistics

For Lloyd's statistics estimation it is assumed that the form of density function for observed values is known, but its parameters (like mean value μ and variance σ) can change. The density distribution function can be determined as following:

$$w(x) = \frac{1}{\sigma} f(\frac{x-\mu}{\sigma}),$$ (1)

If we know the density function (1), we can estimate mathematical expectation and variance of order statistics $\mu = 0, \sigma = 1$:

$$\xi_i = Mx^{(i)},$$ (2)

$$B_{i,j} = M(x^{(i)} - Mx^{(i)})(x^{(j)} - Mx^{(j)}),$$ (3)

Theoretical equations for ξ_i and $B_{i,j}$ values are rather complicated, we do not present them here. For unknown random parameters μ and σ mathematical expectation of order statistics equals to [2]:

$$Mx^{(i)} = \mu + \sigma\xi_i,$$ (4)

where x is a vector of observed order statistics.

For estimation of the unknown parameters it is needed to select values that minimize standard deviation between conditional mathematical expectations (4) and current set of order statistics. Let write simpler form of these parameters for father calculations:

$$A = \begin{bmatrix} 1 & \xi_1 \\ 1 & \xi_2 \\ \vdots & \vdots \\ 1 & \xi_n \end{bmatrix} \text{ and } \Theta = \begin{pmatrix} \mu \\ \sigma \end{pmatrix}.$$

The expression for the least-squares criterion can be written taking into account the correlation of order statistics as: $\Phi = (x - A\Theta)^H B^{-1}(x - A\Theta)$.

Varying the parameter Θ and equating the result to zero, we found the unknown parameters in the following form $\hat{\Theta} = (A^H B^{-1} A)^{-1} A^H B^{-1} x$.

Matrix (5) of found coefficients gives us the required coefficients of linear order statistics [3].

$$(A^H B^{-1} A)^{-1} A^H B^{-1} = \begin{bmatrix} \alpha_1 & \alpha_2 & \cdots & \alpha_n \\ \beta_1 & \beta_2 & \cdots & \beta_n \end{bmatrix},$$ (5)

Mathematical expectation and variance can be found as the following expressions:

$$\hat{\mu} = \sum_{i=1}^{n} \alpha_i x^{(i)},$$ (6)

$$\hat{\sigma} = \sum_{i=1}^{n} \beta_i x^{(i)},$$ (7)

If we do not know the analytic form of distribution, but have a sufficiently large set of empirical data, we can calculate the average value and covariance of order statistics, and then use equations (6) and (7) for finding of approximate values for coefficients α и β.

If $x^{(i)}$ is the set of ranged values, there are some obvious requirements for ordered series which are used to estimate the mean and variance, such as $\mu = f_\mu\left(x^{(1)}, x^{(2)}, \dots x^{(n)}\right)$.

First of all, if we add some constant C to all values, estimations of the mean value should change on the same value. That's why the first condition on the function f:

$$f_\mu\left(x^{(1)} + C, x^{(2)} + C, \dots x^{(n)} + C\right) = f_\mu\left(x^{(1)}, x^{(2)}, \dots x^{(n)}\right) + C , \qquad (8)$$

At the same time, for the function that provides estimation of the variance $\hat{\lambda} = f_\lambda\left(x^{(1)}, x^{(2)}, \dots x^{(n)}\right)$, adding of additional constant C shouldn't change the estimation of the variance. That's why the second condition is:

$$f_\lambda\left(x^{(1)} + C, x^{(2)} + C, \dots x^{(n)} + C\right) = f_\lambda\left(x^{(1)}, x^{(2)}, \dots x^{(n)}\right), \qquad (9)$$

Based on these conditions we can write following equations for coefficients of order statistics:

$$\sum_k \alpha_k = 1, \ \sum_k \beta_k = 0, \qquad (10)$$

3. Nonlinear order statistics filter

Unlike the linear statistics, for neural network with nonlinear activation functions the conditions (8) can't be satisfied, because for some values of constant C activation functions will saturate and properties of estimations will change. Suppose that we know the probability distribution function of the estimated parameters μ и λ, or at least the range of their variations. Then we can create a representative set of training examples, and we need to achieve the results that the neural network gives a good performance for this prior distribution. In this case the exact conditions (8) are not guaranteed, but if the distribution of the estimated parameters corresponds to our expectations, then we will get satisfactory results (figure 1). Design of order statistics based on neural networks has obvious similarity to the estimates based on the criterion of maximum of a posteriori probability.
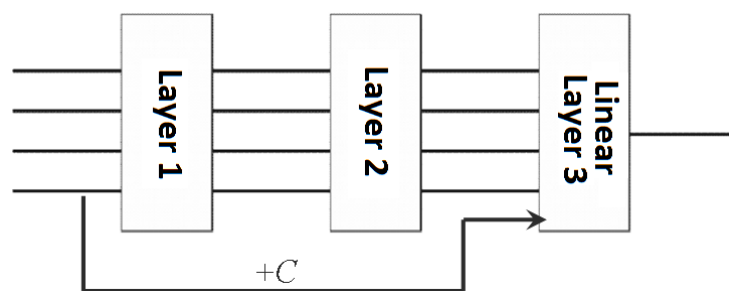

Figure 1. Neural network architecture for order statistics estimations.

It would be better to use information about Lloyd's order statistics for initial distribution of neurons weights [7]. Consider the simplest technique for solving this problem in the case when functions implemented by various network layers are evidently separated. The error that occurs due to the nonlinearity of neurons activation functions prior to the network training should be small. This fact imposes certain restrictions on the choice of activation functions for layers. For example, the hyperbolic tangent (the sigmoid) has a domain with a weak nonlinearity at the origin of coordinates.

While using neurons with sigmoidal activation functions, by properly scaling the training set we can make transformations of a signal by the network neurons close to linear.

Let us consider, as an example, the application of the proposed initialization algorithm to a three-layer feedforward neural network. The input layer of the network scales the data to make all values of the series to be located within the linear domain of the hyperbolic tangent. The intermediate layer performs a weak nonlinear transformation with the help of the weight matrix that contains coefficients of the Lloyd's statistics filter. The output layer performs the inverse data transform. By a proper choice of the α coefficient one can make the error stipulated by the nonlinearity of the activation function arbitrarily small. In this case the neural network with weights chosen in such a way will estimate the current value of order statistics in accordance with expressions (6, 7) [6].

$$
A = \begin{bmatrix} 0 & 0 & 0 & ... & 1/\alpha \end{bmatrix}
\begin{bmatrix}
0 & 1 & 0 & ... & 0 \\
0 & 0 & 1 & ... & 0 \\
... & ... & ... & ... & ... \\
0 & 0 & 0 & ... & 1 \\
-a_1 & -a_2 & -a_3 & ... & -a_p
\end{bmatrix}
\begin{bmatrix}
\alpha & 0 & 0 & ... & 0 \\
0 & \alpha & 0 & ... & 0 \\
0 & 0 & \alpha & ... & 0 \\
0 & 0 & 0 & ... & 0 \\
0 & 0 & 0 & ... & \alpha
\end{bmatrix}
$$

Note that the neural network output before training nearly coincide with the Lloyd's statistics. When training the network with an algorithm with a monotonic decrease of the error (e.g., gradient and quasi-Newton algorithms, conjugate gradients method, etc), the result obtained on any training step is the same or even better than that obtained with the help of the linear statistics. In this paper we use the Levenberg-Marquardt algorithm; it belongs to the class of quasi-Newton methods that guarantes a high convergence rate [1].

4. Results

Consider the synthesis of nonlinear order statistics on a model example with Generalized Extreme Value distribution (GEV) with the parameter k:

$$
w_t(x) = \frac{1}{\sigma} \exp\left(-\left(1 + k\frac{x-\mu}{\sigma}\right)^{-\frac{1}{k}} \right) \left(1 + k\frac{x-\mu}{\sigma}\right)^{-1-\frac{1}{k}}, \tag{11}
$$

This type of distribution is often used to simulate the smallest and largest value among a large set of independent and identically distributed random values representing measurements of some values or observations. The main feature of this distribution is that it is asymmetric (figure 2).
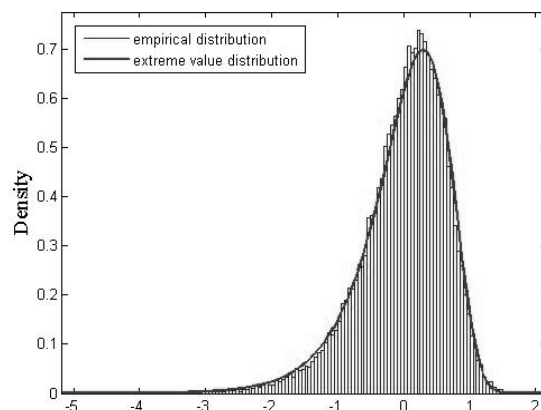

Figure 2. An example of GEV distributed data.

We modeled the set of N samples of the same type, consisting of n counts each. The purpose of this investigation is to compare the Lloyd's statistics coefficients with the neural networks statistics under the assumption that the distribution type is GEV.

The procedure involves the following stages:

1) With the help of a random number generator we create a matrix $N{\times}n$ of random numbers distributed according to the GEV distribution, then we add to the each line a constant C, distributed by the normal law with a mean of 0 and a variance σ_{shift};

2) Sorting the numbers in each lines. Thus, each line forms an ordered sample of n elements, and we have N number of such test samples.

3) Averaging the columns, we find estimates for the expectation of order statistics. Calculating the covariance between the columns, we find an estimation of the covariance matrix for order statistics;

4) According to the equations (4, 5) we calculate two sets of coefficients - for the expectation and variance. After that we can train the neural network. The example of such modelling is shown on the figure 3.
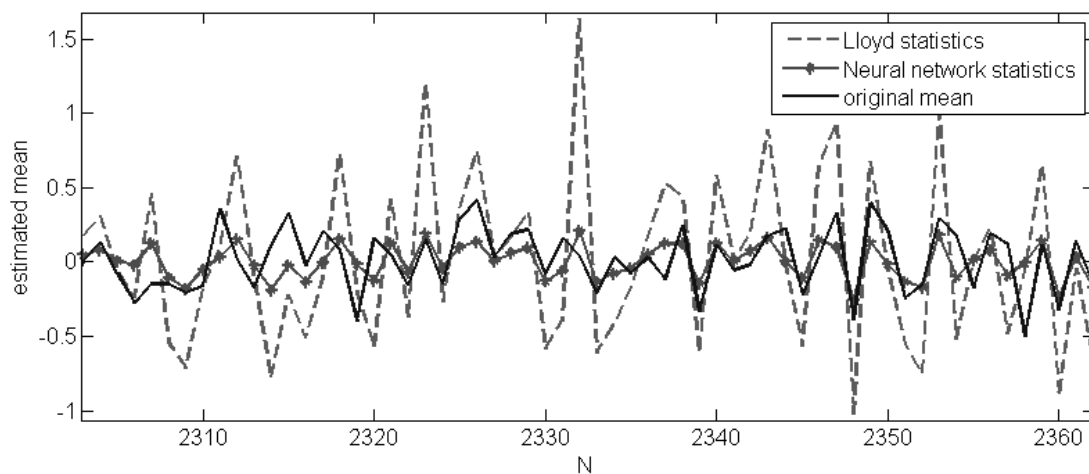

Figure 3. Comparison of expectation values based on Lloyd's and neural network statistics.

Table 1 shows the results of distribution parameters estimations for a sample of random variables with GEV-distribution. The table includes standard deviations for different statistics (Lloyd and neural network statistics) varying the distribution parameter k and the sample size n.

Table 1. Standard deviation of expected value estimations.

Method	n	Parameter k for the GEV-distribution					
		0	0.1	0.5	1	1.5	2
Lloyd's statistics	5	0.479	0.489	0.542	0.610	0.857	0.975
NN statistics	5	0.218	0.217	0.203	0.203	0.823	0.934
Lloyd's statistics	15	0.275	0.274	0.296	0.315	0.334	0.342
NN statistics	15	0.166	0.161	0.167	0.094	0.268	0.214

Figure 4a shows an example of a grayscale image. The same image with GEV-distributed noise is shown on the figure 4b. Figure 4c shows the results of the image filtering based on Lloyds statistics. And the figure 4d shows the results of the neural network statistics filter. It is clearly seen that nonlinear order statistics filter is better that linear order statistics filter.

Figure 4a. Original image.

Figure 4b. Image with a noise of GEV-distribution.

Figure 4c. The result of the Lloyd's statistics filter

Figure 4d. The result of the neural network statistics filter

However, when we work with real data, we cannot determine in advance the type of noise distribution in the processed signal. Therefore, the filtering of a set of real data will include the following stages: in the first stage the separation of noise from the original dataset with the median filtering or moving average, then an assessment of species distribution and its parameters. Then forming a test data with a given type of distribution, and the search for the coefficients of linear and neural network of order statistics, with the help of a network formed by filtering the original data. To assess the quality of filtering, you can repeat the procedure of filtering, starting with the first step.

References
[1]   Haykin S 1999 Neural Networks: A Comprehensive Foundation, *Prentice Hall*
[2]   David H A and Nagaraja H 2003 Order Statistics *Wiley Series in Probability and Statistics*
[3]   Lloyd E H 1952 Least-Squares Estimation of Location and Scale Parameters Using Order Statistics *Biometrika* **39.1/2**
[4]   Chang-Shing Lee, Yau-Hwang Kuo and Pao-Ta Yuhttp 1997 Weighted fuzzy mean filters for image processing *Fuzzy Sets and Systems*. **89** 157-180
[5]   Yiming Pi and Shunji Huang 1995 Design of multistage weighted order statistic filters by a neural network *International Conference on Image Processing (ICIP'95)* **Vol. 1** 346-250.
[6]   Maslennikova Yu S and Bochkarev V V 2011 Nonlinear order statistics using artificial neural network *MEPHI Neiroinformatics-2011 Book of papers* **2** 102-113 (In Russian).
[7]   Maslennikova Yu S and Bochkarev V V 2010 Initialization of multilayer forecasting artifical neural networks *Uchenye Zapiski Kazanskogo Universiteta. Seriya Fiziko-Matematicheskie Nauki,* **152** 7-14 (In Russian). arXiv:1410.6413 (In English)