CASCADE ARTIFICIAL NEURAL NETWORKS

Avshalumov A.Sh., Filaretov G.F.

Moscow Research Institute for Cybernetic Medicine (MRICM)
Moscow Power Engineering Institute (TU)

ABSTRACT

The new variant of artificial neural networks (ANN), named by *cascade ANN* (CANN), is proposed. Such networks are formed as a cascade connection of several artificial neural networks. This type of network is designed for information processing applying to significantly heterogeneous data.

The proposed cascade ANN differs from known variants of hybrid networks by organization of training process: different training samples are used for ANN training of different cascades. These training samples are formed on a basis of results of the ANN training of a previous cascade.

Possibilities of the proposed ANN type are demonstrated in the course of solving following three typical problems: classification with differently remote classes, classification with differently informative data, developing a mathematical model of an object at presence of several operation modes.

Index Terms - Artificial neural networks, cascade ANN, significantly heterogeneous data, classification, model.

1. INTRODUCTION

Artificial neural networks, as practice shows, are sufficiently effective method to solve problems of classification and construction of mathematical models of various objects. However, in some cases, their effectiveness is sharp reduced. Often this is due to the presence of significant structural heterogeneities of the experimental data used for training ANN. In its origin such heterogeneity, as a rule, are caused by specific properties of the phenomenon, changes in the structure of the simulated object or mechanism of the observed process, etc.

To achieve the required efficiency of neural network approach in such situations it is proposed to use combined or cascaded ANN (CANN). Such networks are constructed as a cascade connection of at least two artificial neural networks, which can be of various types, each stage may include several ANN.

In itself, this idea is not new. It is known, for example, the neural network counter-propagation, which is a serial communication of Kohonen network and the perceptron layer [1]. Such networks are sometimes referred to as the hybrid, thus underlining the association of different types of ANN. However, the cascade networks have a number of specific features that distinguish these networks from the hybrid.

2. MAIN FEATURES OF CASCADE ARTIFICIAL NEURAL NETWORKS

Cascade ANN have a special way of their training process organization, which includes the following steps:

- the network of the first cascade is trained using the whole initial training sample;
- training sample(s) for a network (or several networks) of the second cascade is formed according to the training results of the specified network (ANN of the first cascade); such formation can be carried out by sorting initial sample elements and using some conversation of initial information, etc;
- such operations are continued for subsequent cascades if there are any ones or necessity of which is revealed in the course of training.

Thereby, essential distinguishing character of the cascade ANNs is used different training samples for training ANN of different cascades; moreover, these training samples are formed on a basis of results of the ANN training of a previous cascade. In whole the training process itself is realized sequentially from the first cascade to the last.

Consider the possibilities of the CANN for solving concrete tasks mentioned before.

3. CLASSIFICATION WITH DIFFERENTLY REMOTE CLASSES

The task of classification on the basis of usage of artificial neural networks of multilayer perceptron (MLP) type, in conditions when a priori it is known to what exactly class each element $\bar{\mathbf{X}}_g = \left\| x_{g1}, x_{g2}, \ldots, x_{gn} \right\|$, $g = 1, 2, \ldots, N$ of training sample belongs ("supervised learning") is considered. Moreover the possibility of essential heterogeneity in location of points of different classes of the training sample in n – dimensional space is postulated (according to principle "where – thick, where – empty"). In other words, points of some classes are remote from each other on significant distances and so they are easily separated, at the same time points of some other classes are small remote from each other and it is difficult separate them.

Such situation in a simplified form is shown in Figure 1, where the points of classes 1 and 2 are separated easily in contrast to classes 3 and 4; the other classes 3 and 4 together easily separable from each of the classes 1 and 2.

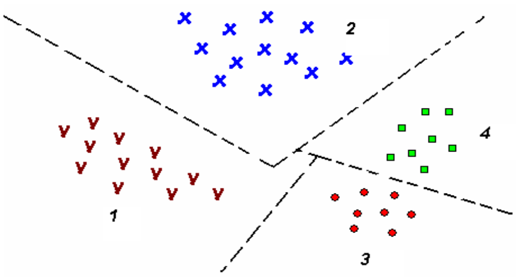

Figure 1: example of differently remote classes

It is possible to name such problem of classification shortly as the classification problem with differently remote classes. The multilayer perceptron in such cases becomes low effective. Then we propose to use cascading several multilayer perceptrons.

Let it is required to distinguish m classes and A_1 , A_2 ,…, A_m – sets of observation points referring to these classes. As a first cascade a multilayer perceptron, containing n inputs and m outputs, is synthesized. If after training such perceptron with acceptable degree of reliability distinguishes all m classes, problem can be considered as solved. However, as a rule, for variant of differently remote classes, the network reliably distinguishes only k classes (for instance, A_1 , A_2 ,…, A_k) and points of the rest classes are badly distinguished, hereupon there is occurrence of significant number of classification errors.

Let's form of these points the following new set $\boldsymbol{B} = A_{k+1} \bigcup A_{k+2} \bigcup \ldots \bigcup A_m$, and covert the initial perceptron for distinguishing points of sets A_1 , A_2 ,…, A_k , $\boldsymbol{B}$. Obviously, it will have $k+1$ outputs. The second cascade is built using new perceptron intended for distinguishing points referring to different classes of set $\boldsymbol{B}$, and having a number of outputs equal to the number of these classes. Herewith only the training sample elements which correspond to the classes of set $\boldsymbol{B}$ are used for its training.

Figure 2 clearly demonstrates the results of the operation of such a two-stage CANN. Perceptron of the first stage classifies the point classes 1, 3 and co-classes 3 and 4, and perceptron of the second stage then divides classes 3 and 4.

It is clear that now all previous considerations connected with the first cascade can be repeated applying to set $\boldsymbol{B}$, etc. Such extension (cascading) of perceptrons lasts until the recognition of all classes has been ensured (certainly, if it is possible). The natural generalization of this consideration is a variant when at the first stage we succeed only to subdivide the initial population A_1 , A_2 ,…, A_k into several subsets B_1 ,B_2 ,…, B_L , $L<m$, each of them possibly contains several elements of A_j .

This situation can also exist at subsequent stages. In any case total amount of the used perceptions will not exceed (m – 1).

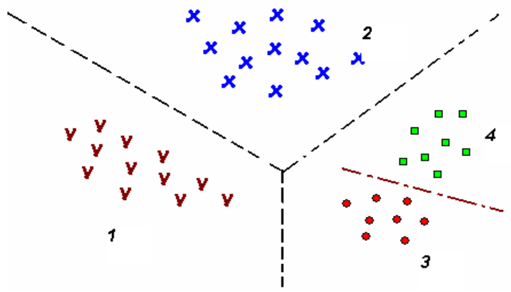

Figure. 2: results of the two-stage CANN operation

At many m this is significantly less than, for instance, in the known method of pairwise classification, when an amount of required perceptrons is equal to $m\,(m-1)/2$.

In fact, this approach also allows to improve significantly the quality of classification in conditions of imbalance of a training sample, when amount of observations referring to various classes differs greatly.

4. CLASSIFICATION WITH DIFFERENTLY INFORMATIVE DATA

The second problem of the classification with differently informative data is also characterized by presence heterogeneity of data forming training sample, but of the other nature. We mean that though the total amount of the observed variables (factors) n can be very high, levels of their self-descriptiveness are often differ. Usually, it is always possible to separate a relatively not numerous group of the most essential factors and the second group (very often very numerous) of low informative variables, however, it should be recklessly to neglect their influence. In case of such different self-descriptiveness of data it is proposed to use their preliminary compression. Within the framework of neural-network paradigm such compression can be implemented by means of autoassociative ANN.

Formally an Autoassociative artificial neural network (AANN) refers to a category of multilayer perceptrons and contains n inputs, the same amount n of outputs, some hidden layers, moreover in one of them – in compressing layer – a number of neurons m n. Values at outputs of neurons of the compressing layer out , out , …, out take initial n – dimensional data into space of smaller dimensionality m practically without loss of main usual information. At the same time this dimensionality decrease can simplify solving the main problem of classification.

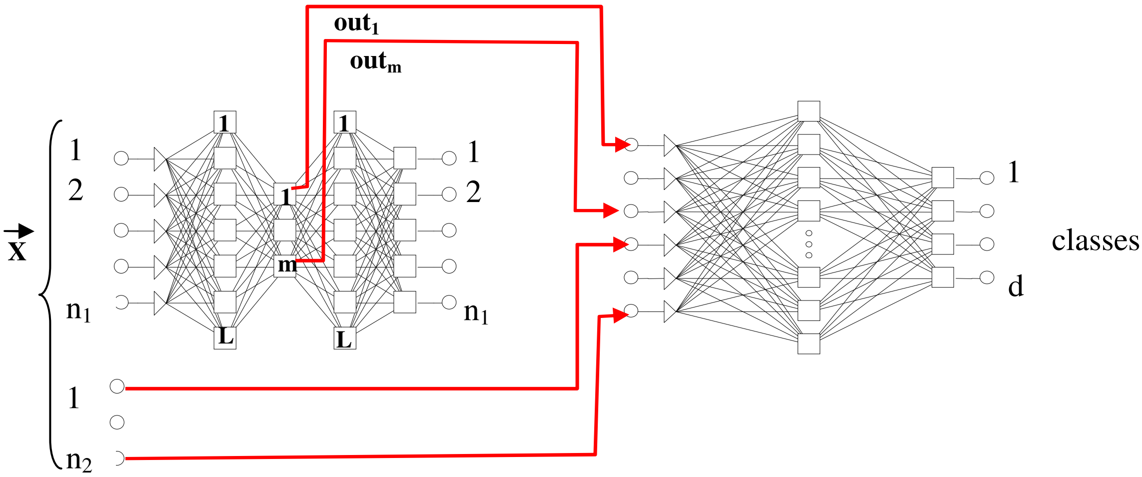

Figure 3: cascaded neural-network with AANN

Within the framework of idea of the neural-network cascading, several variants of ANN can be proposed in which AANN is used as the first cascade, and in the second cascade either self-organizing (self-training) networks (for instance, Kohonen network) or networks using algorithm of supervised learning (for instance, MLP) are used (see Figure 3). This idea can also be used in constructing models of various objects

5. DEVELOPING A MATHEMATICAL MODEL OF AN OBJECT AT PRESENCE OF SEVERAL OPERATION MODES

The solving task is modeling static nonlinear object with varying characteristics. It is assumed that output of object ***Y*** depends on a vector of input controllable and measured factors ***X*** and on a vector of measured but not controllable variables ***Z***; besides, there is a random additive noise *e* distorting output measured values (Figure 4).

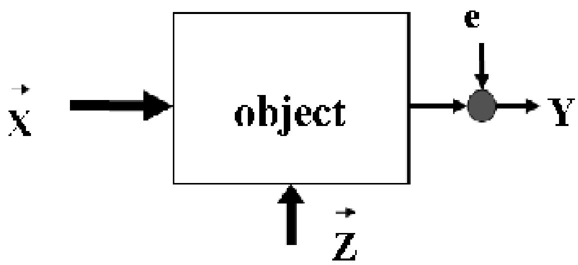

Figure 4: static nonlinear object

In contrast to standard statement of the problem the situation when an object can function in different models determined by correlations between separate components of vector ***Z,*** is considered. In this case it is often difficult to detect a moment of conversion from one mode into another – for instance, due to natural fluctuation of characteristics of used raw material.

Suppose that at the present object has *d* different modes of operation $U_1(\vec{\mathbf{Z}})$, $U_2(\vec{\mathbf{Z}})$, …, $U_d(\vec{\mathbf{Z}})$ and, moreover, the establishment of the moments of regime change is very difficult.

Each mode can have its response function: $\varphi_1\left(\vec{\mathbf{X}},\vec{\mathbf{Z}},U_1(\vec{\mathbf{Z}})\right),\ldots,\varphi_d\left(\vec{\mathbf{X}},\vec{\mathbf{Z}},U_d(\vec{\mathbf{Z}})\right)$. In such situation, application of traditional indentification procedures, at its core, allows to obtain the response function only averaged on all set of modes that make it practically unusable, for instance, for optimization goals of technological process.

In fact here it is necessary to solve two interconnected problems. First of them is a problem of recognition (classification) of a mode; second one is a problem of estimation of response function on a structured sample when estimating function (***X*** , ***Z***) is carried out with using only those elements of the sample which are referred to the given j-mode.

The cascade networks are very well adapted for identification of such type objects. During solving the first problem any one of ANN can be used, intended for classification goals using the method of unsupervised learning (for instance, self-organizing Kohonen network). The network of such type forms the first cascade of CANN. For solving the second problem it is reasonable to apply two-layer perceptrons. The total amount of them will be equal to a number of different models, and the whole population integrally is the second cascade of the network.

6. EXAMPLES OF CANN PRACTICAL USE

6.1. Differently remote classes

Cascaded Neural-network was applied in medical complex DKM-01 [2]. This complex is intended for integral estimation of functional condition of human separate organs and tissues according to the following gradations: "norm" is class 0, "satisfactorily" is class 1, "unexpressed pathology" is class 2, "expressed pathology" is class 3. As input information the set of 11 the most informative parameters, obtained during pre-processing of the original signal, recorded during diagnostic procedures, is used.

As classify ANN was initially selected three-layer perceptron with a number of inputs n = 11 and the number of outputs m = 4 (the number of classes). But all attempts to create ANN using known approaches turned out to be not very successful because it became clear that classes 0 and 1 differ very little. Practically all events of class 0 were classified as class 1. Objectively, most likely this is connected with the following: differences between gradations “norm” and “satisfactorily” are often difficult to catch.

In this connection it was solved to use the idea of cascading networks. As the first cascade a three-layer perceptron was used with number of inputs n = 11, three inputs (m = 3). It was trained to distinguish classes 2 and 3, as well as integrated class B made up of events referring to initial classes 0 and 1.

The second cascade also represents a perceptron with eleven inputs, two hidden layers and two outputs. For training this network only those elements of training sample was used, which were referred by the perceptron of the fist cascade to class B.

With help of the trained network testing it was established that the correct classification occurred more than 80% of cases. Taking into account the very complex nature of the problem, learning results can be considered more than satisfactory.

6.2. Differently informative data

The problem of the building forecasting model of random content in atmosphere of workings of mine, according to experimental data was solved. It is supposed that main factors affecting radiation situation of mine atmosphere are the following mining and technological parameters:

- X_1 is a chamber volume, (m^3);
- X_2 is part of chamber volume which fall at volume of panel (section);
- X_3 is average daily mass of blown explosive (kg);
- X_4 is a part of chamber volume corresponding to a unit of mass of the blown explosive (m^3/kg);
- X_5 is a type of ventilation branch in place of sampling (X_5 = 1 if a ventilation work of incoming air stream, X_5 = 2 if a ventilation work of operating air stream, X_5 = 3 if a ventilation work of outcoming air stream);
- X_6 is an amount of aerodynamic bonds of the panel (section);
- X_7 – fissuring (depending on degree of fissuring of variable X_7 (X_7 = 3 – fissuring is strong, X_7 = 2 – fissuring is average, X_7 = 1 – fissuring is weak);
- X_8 is a depth of panel (chamber) processing, (m);
- X_9 is availability of ventilator of local venting (VLV) in the chamber (at availability of VLV - X_9 = 1, at absence - X_9 = 0);
- X_{10} is ore production per shift , (t).

Value of radon concentrations (*Mel/l*)/1000 is used as response Y. It allows to evaluate a radon content at the average over the mine and, when knowing volumes of outgoing air, to estimate effect of the mine by radiation factor to ecology of atmosphere in its location.

The experimental data obtained by radiation safety service and containing 31 observations were used for building this model. The above experimental data were earlier used for building a regression model. As a result it was obtained adequate linear model containing only three significant factors X_2, X_4, X_8 : Y= 24,7971 – 12,7848 X_2 +0.0120 X_4 -0.0716 X_8 with residual variance σ^2_{res} = 8,037.

The different variants of ANN for obtaining the similar model were studied. It was shown that the best variant is obtained when the cascade network is used. Practicability of using cascade network is connected with the circumstance that a part of factors in this problem carries a quantitative nature and a part has a quantitative nature or a nature close to it. Factors X_1, X_3, X_5, X_6, X_7, X_9 can be referred to the last category. So it is proposed to apply the cascade network where autoassociative network $I_5h_7h_2h_7O_5$ is used as a first cascade and two-layer perseptron $I_7h_3O_1$ is used as a second cascade.

The first network is trained for reproduction of values at its output, and after that values S_1, S_2 are fixed generating at output of compressing layer h_2. For training the second cascade perseptron a new data set is generated including 31 observations of factors X_1, X_2, X_4, X_8, X_{10} , ,appropriate values of variables S_1, S_2 and output Y. As a result the model was obtained with the following quality indexes: variance of prediction error for elements of the training sample (21 observations) turned out to be equal to: σ^2_L = 2.36, for elements of the test sample (10 observations) σ^2_T = 7.58, for all observations: σ^2_N = 3.89.

The analysis of the obtained neural-network model indicates that the partial dependence $Y = \varphi_1(X_2, X_4)$ is really close to linear (see Figure 5).

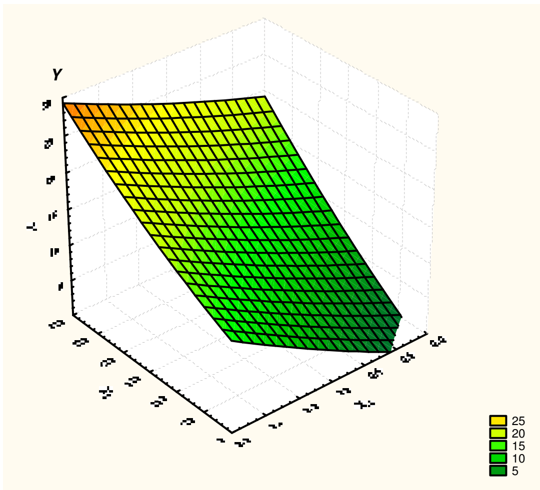

Figure 5: the partial dependence $Y = \varphi_1(X_2, X_4)$

However, influence of variable X_8 is clearly non-linear (see, for example, Figure 6).

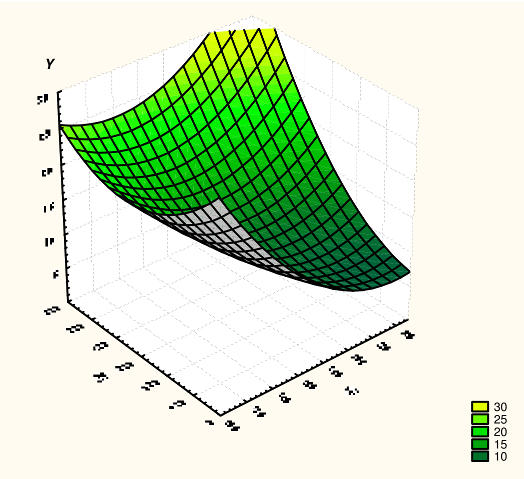

Figure 6: the partial dependence $Y = \varphi_2(X_4, X_8)$

Likewise, the dependence $\varphi_4(S_1, S_2)$ is also nonlinear (see Figure 7).

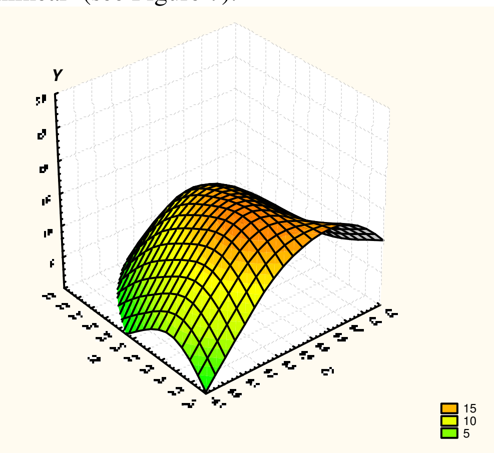

Figure 7: the partial dependence $\varphi_4(S_1, S_2)$

Obviously, it is also impossible to neglect total influence of factors of qualitative character X_3, X_5, X_6, X_7, X_9 .

6.3. Objects with several operation modes

The proposed ANN type was successfully tested on example of building model of a pilot plant for chemical and pharmaceutical purpose. This plant has two controllable factors X_1 and X_2 and only three measured variables Z_1, Z_2 and Z_3 (moisture of raw material, crushing and properties of the filler). It is known that, depending on the variables Z_1, Z_2, Z_3 plant can operate in two different modes.

It is necessary to optimize settings for each of the modes, implementing a procedure to maximize the productivity indicator Y (while ensuring proper quality of products) on the basis of the static models. It is believed that the properties set on the slow change in time, as permitted above values of control inputs X_1 and X_2 are in the range (0 - 1).

For solving this task the two stage network was used: the first stage - self-organizing Kohonen network with three inputs and two outputs, the second stage – two perceptrons $I_5h_{10}O_1$.

The sample for the identification contained N = 1011 observations, from which $N_O = 911$ was intended for training, and the remaining N_T = 100 (selected randomly) - for testing purposes. As a result of the classification with help of Kohonen network all elements of the original sample is divided into two groups, designed to build identification models $Y = \varphi_1(X_1, X_2)$ and $Y = \varphi_2(X_1, X_2)$ for each of the two identified modes. These models were built with the help of two perceptrons of the second stage. These neural network approximations of dependences of the output indicator Y from controllable factors X_1 and X_2 (with fluctuations of the measured variables Z_1, Z_2 and Z_3) are shown in graphic form in Figures 8 and 9.

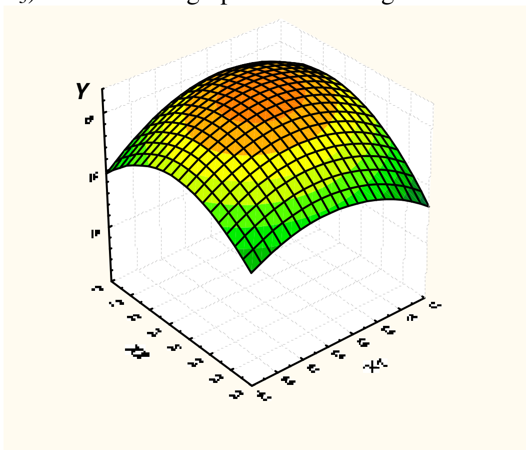

Figure 8: the dependence $Y = \varphi_1(X_1, X_2)$

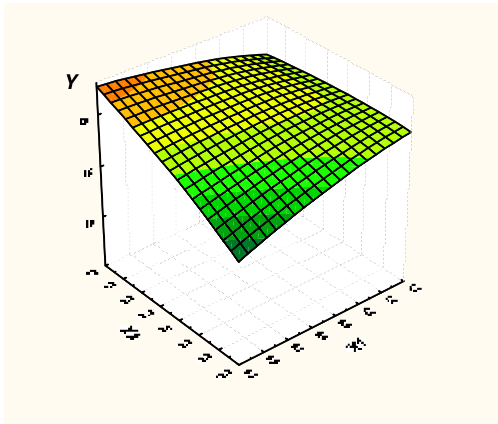

Figure 9: the dependence $Y = \varphi_2(X_1, X_2)$

Obviously, there is a significant difference in the response function for the first and second modes.

7. REFERENCES

[1] Simon Haykin, Neural Networks. A Comprehensive Foundation. Second Edition. Prentice Hall Inc, 1999.

[2] Avshalumov A.Sh., Sudakov K.V., Filaretov G.F. New information technology for system diagnosis of functional activity in human organs. Medical Technology, 2006, № 3, pp. 13 – 18 (rus).