University of Ostrava

Institute for Research and Applications of Fuzzy Modeling

Backpropagation and his application in ECG classification

Ondřej Polakovič

Research report No. 75

2005

Submitted/to appear:

Journal of Electrical Engineering

Supported by:

1M6798555601 of the MŠMT ČR

University of Ostrava
Institute for Research and Applications of Fuzzy Modeling
Bráfova 7, 701 03 Ostrava 1, Czech Republic

tel.: +420-69-6160234 fax: +420-69-6120 478
e-mail: e-mail ondrej.polakovic@osu.cz

We show on an example from medical diagnosis that some problems can be solved using simple neural networks. First we define some basic notions from neural network theory. We mention also some basic facts about electrocardiography. Then we use three-layered neural network with backpropagation algorithm to adaptation on classification the patients' ECG signals into two classes and summarize results.
Keywords Neural Networks, Backpropagation, ECG,Learning and adaptive systems

2000 Mathematics Subject Classification: 82C32

1 INTRODUCTION

Artificial neural networks [9, 4, 2] can be viewed as simplified mathematical models of brain-like systems and their functions as parallel distributed computing networks. The study has its roots in the work of McCulloch and Pitts [8] and Hebb [6].

Perhaps the most important advantage of neural networks is their adaptivity. Neural networks can adjust the weights to optimize their behavior as pattern recognizers, decision makers, system controllers, predictors, etc. While fuzzy logic performs an inference mechanism under cognitive uncertainty, neural networks offer advantages, such as learning, adaptation, fault-tolerance, parallelism and generalization. However, it is not possible to extract rules from trained neural network (black-box), and we can not generally integrate special information about the problem into the neural network. A certain input produces a desired output, but how the network achieves this result is left to a self-organizing process. The neural networks can be divided (according to the topology) into two groups - recurrent and feed-forward or can be divided (according to the learning algorithm which adapts the weights) to neural networks with supervised learning (with teacher) or with unsupervised learning.

In this report we use the backpropagation algorithm to adapte the neural network. The problem is to classify the patients' ECG to class of normal ECG and abnormal ECG. We show our successful results of adaptation and classification

This research report is organized as follows: in Section 2, we mention some basic notion from the neural network theory, in Section 3, we define backpropagation algorithm, in Section 4, we describe basic principle of ECG. In Section 5, we show one application from medicine and in the last Section 6 there is our conclusion about results obtained from adapted network.

2 DEFINITIONS

The basic processing units of neural networks are called *artificial neurons,* (or simply *neurons* . See Fig.1). The signal flows from neuron inputs $x_i,\ i = 1, \cdots, n$, then interacts with the weights $w_i\ i = 1, \ldots, n$ to produce product $p_i = w_i x_i\ i = 1, \ldots, n$. The input information p_i is aggregated (by addition) to produce input to the neuron

$$net = w_1 x_1 + \ldots + w_n x_n. \tag{1}$$

The neuron uses its *transfer–activation function* $f : R \longrightarrow R$ to compute the output

$$y = f(net) = f(w_1 x_1 + \cdots + w_n x_n). \tag{2}$$

The so-called logistic function

$$f(net) = \frac{1}{1 + e^{-\beta net}} \tag{3}$$

is often used. However, we can also use linear, tangential, sigmoidal functions, etc.

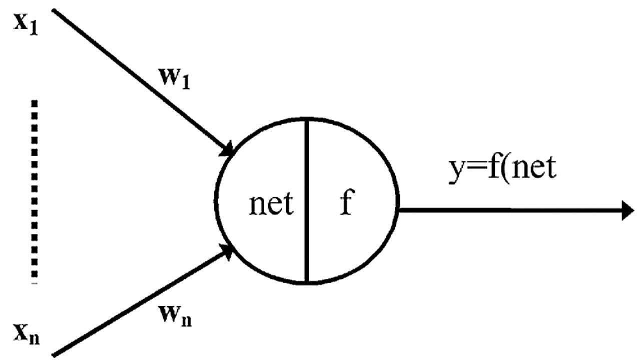

Fig.1. Structure of artificial neuron.

We will denote number of elements from the set M as $|M|$.
Below is the classical formal definition of the neural network.

Definition 1 *A neural network is a tuple* (U, W, A, O, NET, ex)*, where:*

1. U *is a finite set of processing units (neurons).*
2. $W : U \times U \longrightarrow \mathbb{R}$*, is a function representing the network structure.*
3. $A : U \longrightarrow \mathbb{R}^{\mathbb{R}}$ *assigns an activation function to each* $u \in U$*.*
4. $O : U \longrightarrow \mathbb{R}^{\mathbb{R}}$ *assigns an output function to each* $u \in U$*.*
5. $NET : U \longrightarrow \mathbb{R}^{(\mathbb{R}\times\mathbb{R})^{|U_i|}}$ *assigns a networks input (propagation) function to each* $u \in U$ *except of input layer, where* U_i *is the set of the neurons connected into the neuron* u*.*
6. $ex : U' \longrightarrow \mathbb{R}$ *is an external input function (assigns an external input to each* $u \in U'$ *where* U' *is the input layer).*

The multilayer perceptron is an extension of the simple perceptron introduces by Rosenblatt in [10]. It is a feedforward neural network that consist of an input layer, one (or multiple) hidden layer and an output layer.

Definition 2 *A multilayer perceptron is a neural network* (U, W, A, O, NET, ex) *that has following properties :*

1. $U = U_1 \cup \cdots \cup U_n$ *is a set of processing units (neurons) where* $n \geq 3$*. Furthermore,* $U_i \neq \emptyset, i = 1, \ldots, n$ *and* $U_i \cap U_j = \emptyset$ *for* $i \neq j$*.* U_1 *is called the input layer and* U_n *is called the output layer. Between are hidden layers.*
2. *The network structure is given by the function* $W : U \times U \rightarrow \mathbb{R}$*, where only connections between consecutive layers are supposed.*
3. $A : U \rightarrow [0, 1]^{\mathbb{R}}$ *assigns an activation function* $A(u) = a_u$ *to each unit* $u \in U$ *where each function* a_u *actives neuron* $u \in U$ *as follows:*
$$a_u = ex(u) \tag{4}$$
for all $u \in U_1$*, and*
$$a_u = f(net_u) \tag{5}$$
for all $u \in U_i$*,* $i \in \{2, \ldots n\}$*, We suppose that all units use the same non-linear function* $f : \mathbb{R} \longrightarrow [0, 1]$.
4. $O : U \longrightarrow [0, 1]^{\mathbb{R}}$ *assigns an output function* $O(u) = o_u$ *to each unit* $u \in U$*. In our case we suppose that* $o_u = a_u, u \in U$*.*

5. *$NET : U \longrightarrow \mathbb{R}^{(\mathbb{R}\times\mathbb{R})^{|U_{i-1}|}}$ assigns a network input (propagation) function to each unit $v \in U_i$ $(2 \leq i \leq n)$ to compute the network input net_v with*

$$net_v = \sum_{u \in U_{i-1}} W(u,v) \cdot o_u + \theta_v \tag{6}$$

$\theta_v \in \mathbb{R}$ is the bias (threshold) of unit v.

6. *$ex : U_1 \longrightarrow [0,1]$ assigns an external input $ex_u = ex(u)$ to each input unit $u \in U$.*

The activation of the neurons is taken from [-1,1] or any other real interval.

3 BACKPROPAGATION

The learning algorithm for multilayer perceptrons requires a differentiable activation function, frequently is logistic function used(non-linear, monotonic, increasing, differentiable). The term *backpropagation* means the backward propagation of an error signal through the network. After propagating a pattern through the network - *feedforward,* the output pattern is compared with a given target and the error of each output unit is calculated. This error is propagated backwards to the input layer - *backpropagation.* Finally the errors of the units are used to modify the weights.

This rule introduced by Widrow and Hoff is also called *generalized delta rule.*

Let have the learning problem L, which is represented by set of pairs of vectors in form $\{i^{(p)}, t^{(p)}\}$, where $i^{(p)}, t^{(p)}, p = 1, \ldots, k$ represents m-ary vectors of learning input patterns and k-ary vectors of learning output patterns, resp.

Definition 3 *Let a multilayer perceptron with $U = U_1 \cup \ldots \cup U_n$will be given using a sigmoid activation function $f(net_u)$ for all $u \in U$ and let L be a fixed learning problem. The supervised learning algorithm that determines the modifications for the networks structure W after propagation is given by*

$$\Delta_p W(u,v) = \eta \delta_v^{(p)} a_u^{(p)} \tag{7}$$

with $u \in U_i,\ v \in U_{i+1},\ i = 1, \ldots, n-1$ and $\eta > 0$, where

$$\delta_v^{(p)} = f'(net_v^{(p)}) \sum_{v* \in U_{j+1}} \delta_{v*}^{(p)} W(v, v*) \tag{8}$$

if $v \in U_j, j = 2, \ldots, n-1$ or

$$\delta_v^{(p)} = f'(net_v^{(p)})(t_v^{(p)} - a_v^{(p)}) \tag{9}$$

if $v \in U_n$. The $p \in L$, $a_u^{(p)}$ is the activation of unit u after propagation of input pattern $i^{(p)}$ and $t_v^{(p)}$ is the target output of unit $v \in U_n$.

The method introduced in Definition 3 is called the *generalized delta rule or backpropagation* algorithm because the backpropagation algorithm implements a *gradient descent method.* The goal of the learning is to minimize the error

$$E = \sum_{p \in L} E^{(p)} = \frac{1}{2} \sum_{p \in L} \sum_{v \in U_n} (t_v^{(p)} - a_v^{(p)})^2 \tag{10}$$

where for the weight changes we assume

$$\Delta_p W(u,v) = -\frac{\partial E^{(p)}}{\partial W(u,v)}. \tag{11}$$

The determination of the weight changes usually does not follow exactly the procedure outlined in the derivation of the learning process. If the last weight change was triggered by the propagation of the pattern q then the new weight change caused by the next pattern p is

$$\Delta_p W(u,v) = \eta \delta_v^{(p)} a_u^{(p)} + \beta \Delta_q W(u,v) \tag{12}$$

where the parameter (from $\mathbb{R}$)$\eta > 0$ is a *learning rate* and β is a *momentum.*

The determination of these parameters depends heavily on factors such as learning problem and initial weights. The values are usually chosen from interval [0,1]. (In our case we fix $\beta = 0$ and try to experiment with $\eta = 0.5, 0.6, 0.7, 0.8, 0.9, 1$ but no obvious differences was found out.)

4 ECG, PQRST COMPLEX

ECG (EKG in some countries) stands for *electrocardiogram,* or *electrocardiograph.* The ECG can provide evidence to support a diagnosis, and in some cases it is crucial for patient management. The contraction of any muscle (and heart of course) is associated with electrical changes called *depolarization* and these changes can be detected by electrodes attached to the surface of the body and the ECG monitor.

The basic shape of normal ECG is shown on fig. 1. The muscle mass of the atria is small compared with that of the ventricles, and the electrical change accompanying the contraction of the atria is small. Contraction of the atria is associated with the ECG wave called P. The ventricular mass is large, and so there is a large deflection of the ECG, when the ventrix are depolarized. This is called the QRS complex. The T wave is associated with the return of ventricular mass to its resting electrical state (repolarization).

The description of human expert — a doctor should always be given in the same sequence. The interpretation indicates whether the record is normal or abnormal. If abnormal, the underlaying pathology needs to be identified. More facts and information can reader find in a plenty of books or articles (e.g. [5, 11, 12]).

Standard expert systems are possible to solve diagnostic of disorder or defect of human's heart but using thousands IF-THEN rules, so it is easy to see, that it is very complicated to obtain the relevant set of rules from human expert to classify patients correctly.

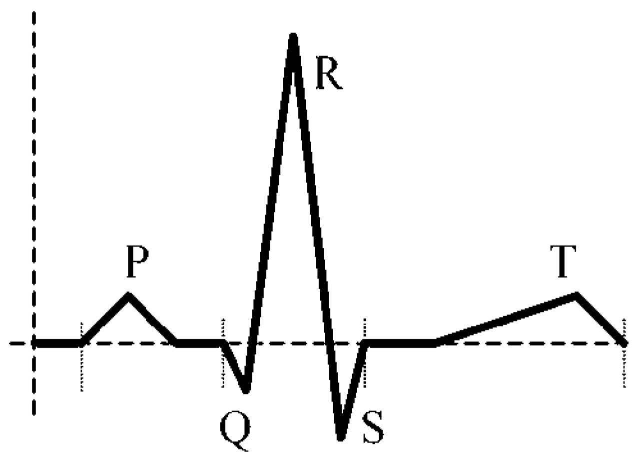

Fig.1. Schematic graph of normal ECG

5 APPLICATION

In this section we will apply the approach described in [7, 3]. We adapt the three-layered neural network (with 100 input neurons, variable number of hidden neurons and one output neuron) using backpopagation algorithm to obtain correct answer to concrete pattern - discretized values of ECG from some patients. The main problem is finding the matching learning set of patterns of ECG signals and setting proper topology of the neural networks. We show on the data obtained in the University Hospital Ostrava–Poruba that it is possible to classify patients into two classes. First with correct — normal ECG and the second group of diseased patients (in this case before the surgery in the hospital).

In cooperation with the University Hospital Ostrava– Poruba we have obtained 36 and 34 ECG signals of wealthy and diseased patients in all. The signals was equidistantly discretized to 100 values from interval [-1,1] for each patient's pattern. We used portable ECG monitor, A/D convertor ADDA Junior and a notebook.

The training set consists of 20 + 20 patterns randomly selected from all patterns. We have assigned eight various training sets to check that it is necessary to set matching learning set.

The results were verified on the learning set and on the rest of the patterns (70 in total) obtained in hospital (See tab. 1, 2, 3). Other learning sets (4, ..., 8) give similar results. Tables 1, 2, 3 represents first, second and third choice of learning set, resp.

Topology	Success.	Percent
100-25-1	68	97%
100-50-1	67	96%
100-75-1	67	96%
100-100-1	68	97%

Tab.1. Results with first learning set.

Topology	Success.	Percent
100-25-1	66	94%
100-50-1	69	99%
100-75-1	68	97%
100-100-1	67	96%

Tab.2. Results with second learning set.

Topology	Success.	Percent
100-25-1	68	97%
100-50-1	70	100%
100-75-1	70	100%
100-100-1	70	100%

Tab.3. Results with third learning set.

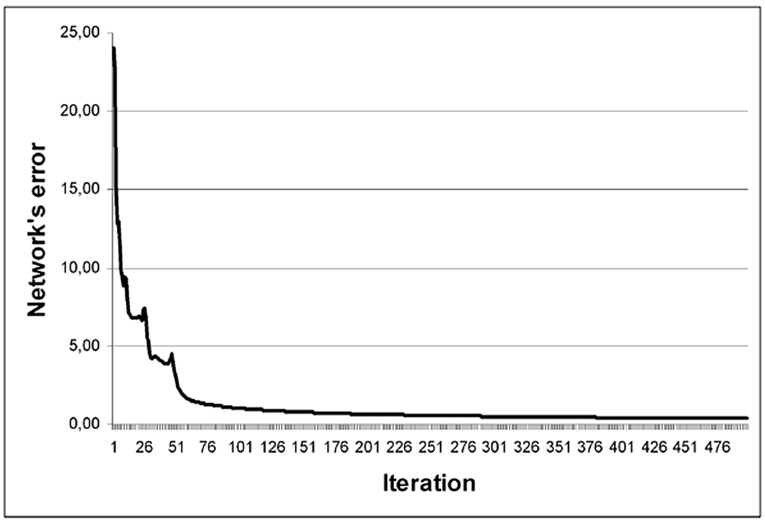

Fig.2. Graph of network error in 500 iteration.

6 Conclusion

In this article about neural networks we have shown one application from medical area. Using three-layered neural networks with sigmoidal - logistic activation function and standard backpropagation algorithm we adapted the network successfully to problem of setting correct class of patients. From about 100 iterations the networks is adapted and classify correctly (for given learning set) and error of network was relatively small and there was no sense to adapt network more than 1000 iterations (See fig.2). But good results depends on choosing the correct learning set - good representing patterns for adaptation.

Acknowledgement This investigation has been partially supported by projects 1M6798555601 of the MŠMT ČR.

References

[1] Abe, Shigeo.: Pattern Classification Neuro-fuzzy Methods and Their Comparison. Springer – Verlag, London 2001.

[2] Bělohlávek, R.: Networks Processing Infeterminacy. Doctoral Thesis. VSB – Technical University Ostrava 1998

[3] Bogacz, R., Markowska-Kaczmar, U., Kozik, A. Blinking artefact recognition in EEG signal using neural network in Fourth Conference Neural networks and Their Applications. Czestochowa 1999.

[4] Fullér, R.: Introduction to Neuro-Fuzzy Systems. Physica–Verlag, a Springer–Verlag Company, 1999.

[5] Hampton, J.R.: EKG short, clearly and comprehensivly. Grada Publishing 1996. (in Czech)

[6] Hebb, D.O.: The Organization of Behavior. Wiley, New York 1949

[7] Jaku. V.: The concept and aplications of artificial neural networks in medicine. Bratislavske Lekarske Listy, 1999; n. 11. 625-637 (in Slovak)

[8] McCulloch, W.S and Pitts,W.: A Logical Calculus of the Ideas Immanent in Nervous Activity. Bulletin of Mathematical Biophysics, 5:115-133

[9] Rojas,R.: Neural Networks. A Systematic Introduction. Springer–Verlag Berlin Heidelberg, 1996.

[10] Rosenblatt,F.: The perceptron: A Probabilistic Model for Information Storage and Organization in the Brain. Psychological Review 65:386-408 1958

[11] Schröder, R. Südhof, H.: The evaluation of EKG in practise. Avicenum Praha 1973 (in czech)

[12] Zeman, K.: The failures of heart's pulsation in an intensive care. Institut pro další vzdělávání pracovníku ve zdravotnictví 1996. (in Czech)

Ondřej Polakovič(Mgr.) is a postgraduate student of Applied mathematics - Fuzzy modeling at the Faculty of Science of the University of Ostrava. His supervisor is Professor Vilém Novák.