PERIODICA POLYTECHNICA SER. EL. ENG. VOL. 42, NO. 1, PP. 103–114 (1998)

NOVEL OPTIMIZATION TECHNIQUES BY NEURAL NETWORKS

János LEVENDOVSZKY

RTD USA BME Laboratories
Department of Telecommunications
Technical University of Budapest
H–1521 Budapest, Hungary
Tel: 36 1 463 3547
Fax: 36 1 463 2485
e-mail: levndov@hit.bme.hu

Received: Oct. 20, 1997

Abstract

Optimization plays a significant role in almost every field of applied sciences (e.g., signal processing, optimum resource management, ...etc.). In spite of the ever-growing need for implementable solutions, the traditional methods of optimization suffer from the drawbacks of either failing to achieve the global optimum or yielding high complexity algorithms which are numerically cumbersome to perform. As a result, novel techniques using neural networks (e.g. Boltzmann machines and Hopfield networks) have been instrumental to optimization theory. Unfortunately, these novel techniques fell short of the expectations for two reasons: (i) in the case of statistical optimization the associated computational complexity is rather high leading to tedious algorithms, and (ii) in the case of Hopfield network only local optimization can be carried out. Therefore, the aim of this paper is to introduce new algorithms for global minimization of Multi-Variable Quadratic Forms (MVQF) defined over discrete sets and for the statistical resource management problem.

The global minimization of MVQFs will be solved by a modified Hopfield algorithm, where the convergence speed is proven to be a polynomial function of the dimension (in contrast to the exponential complexity of exhaustive search). A straightforward application of the result is to implement low complexity algorithms for the detection problem of linearly distorted signals corrupted by Gaussian noise.

The constrained optimization problem of statistical resource management will be interpreted as a set separation problem approximated by a neural network. Based on the underlying tail estimation of the aggregate load, the weights of the network can be properly trained. The results can be directly applied to the traffic design of communication networks, automated factories, ...etc.

Keywords: quadratic optimization, tail estimation, resource management.

1. Introduction

An MVQF is defined as $Q(\overline{y}) = \overline{y}^T\overline{W}\overline{y} - 2\overline{b}^T\overline{y}$, where the global minimum $\hat{y}$: $\text{glob}\min_{\overline{y}\in\{-1,1\}^N} \overline{y}^T\overline{W}\overline{y} - 2\overline{b}^T\overline{y}$ is sought over the finite set of all N dimensional binary vectors. This problem frequently occurs in signal processing and communication theory. For example, the optimal Bayesian detection under Gaussian noise over a linearly distorted channel reduces to the global minimization of an MVQF (see [10,12,13]) and designing structures termed as associative or content addressable memories also involve MVQFs [14].

In the case of discrete sets the minimization does not lend itself to analytical solutions because no gradient can be established. The traditional approach to the discrete problem is the so-called exhaustive search, which in turn yields exponentially growing complexity with respect to the dimension. In order to minimize MVQFs without significant computational overhead, new nonlinear methods were developed (like HOPFIELD net in [1,2]), for which the Lyapunov function was proven to be quadratic. This assures the convergence of the underlying algorithm to the extremum of the quadratic form. The shortcoming of this method results from the complex topology introduced by the optimization algorithm among the elements of $y \in Y$ (often referred to as states). This topology can give rise to several local minima what prevents to capture the global minimum. To cope with this difficulty, statistical optimization methods (e.g. simulated annealing, Boltzmann machines [9]) came into use introducing large computational overhead. As a result, the question of developing fast and global optimization algorithms for MVQFs based on neural architectures remained opened.

Optimal resource management is a central problem of multi-access networks (e.g. ATM networks) connecting random sources together [6,7]. The problem can be modelled as having a user population denoted by $1, \ldots, J$, while $X_j(t)$ random process refers to the random load presented by the jth user. The aggregate traffic is expressed as $Y(t) = \sum_j X_j(t)$, which is compared with the system capacity C. Congestion or overload occurs when $\sum_{j=1}^{J} X_j(t) > C$. The probability of this event should be kept under a certain threshold dictated by the $Q \circ S$ parameter γ, according to the following inequalities:

$$\sup_t P\left(\sum_{j=1}^{J} X_j(t) > C\right) < e^{-\gamma} \quad \text{or} \quad P\left(\lim_{t\to\infty} \sum_{j=1}^{J} X_j(t) > C\right) < e^{-\gamma} . \qquad (1)$$

The task of the resource manager is to enforce inequality (1) by controlling the number of sources. It is obvious that this task comes down to the tail estimation of the aggregate load. As the tail does not lend itself to analytical evaluation, the central problem is to develop an efficient tail estimator with the following properties:

- the estimator is computationally simple for performing real time management function;
- simple descriptors are required from the users which characterize their load (first or second order statistics, without estimating their probability distribution);

- despite the computational simplicity and weak description of sources a sharp estimation is to be achieved.

This casts statistical resource management as a constrained optimization problem, where neural networks can be of help.

The aim of the paper is to introduce neural based methods for solving the above detailed problems by using a modified Hopfield net for quadratic optimization and the set separation approach for optimal resource management.

2. Modified Hopfield Net for Quadratic Optimization

The original Hopfield net [1] is given by the following algorithm with a sequential updating rule.

$$y_i(k+1) = \operatorname{sgn}\left\{\sum_{j=1}^{N} W_{ij} y_j(k) - b_i\right\} . \tag{2}$$

The Lyapunov function of this algorithm is quadratic which implies convergence to the extremum of $Q(\overline{y}) = \overline{y}^T \overline{W} \overline{y} - 2\overline{b}^T \overline{y}$. Two shortcomings occur in this solution, however:

1. the algorithm can get stuck in one of the local optima instead of achieving the global optimum;
2. only maximization of positive definite quadratic forms can be accomplished, though many applications require minimization (e.g. nearest neighbour type of tasks in detection and recognition theory).

To overcome these difficulties the following algorithm is proposed [3,4]:

$$y_i(k+1) = -\operatorname{sght}_{r_i}\left\{\sum_{j=1}^{N} W_{ij} y_i(k) - b_i\right\} \tag{3}$$

which can be rewritten in the form of $y_i(k+1) = -\operatorname{sgn}\left\{ \sum_{j=1}^{N} W_{ij} y_i(k) - b_i - \right.$ $\left. - r_i y_i(k) \right\}$.

The novelty lies in the negative hysteretic type of nonlinearity. While the negative sign assures the minimization, the hysteresis with an appropriately chosen width parameter (r_i) enforces that the algorithm converges to only one steady state corresponding to the global minimum of the underlying quadratic form. More precisely the main result can be summarized in the following theorem:

THEOREM 1 *If*

(i) $\mathbf{W}$ *is a symmetric matrix which is eye-opened with parameter* D *and has positive diagonal elements,*

(ii) there exists an $\mathbf{m}$ *such that* $\mathbf{Wm} = \mathbf{b}$ *and* $m \in dC$*, where*

$$dC := \{u : \varepsilon \leq |u_i| \leq 2 - \varepsilon\}, \tag{4}$$

$$\varepsilon = \frac{3 + k/\min_i \sum_{j,j\neq i} |W_{ij}|}{1 + D}, \tag{5}$$

and

(iii) the hysteresis parameter r_i *is defined by*

$$r_i = W_{ii} + k \quad \text{for some } k > 0,$$

then (i) algorithm (3) has one and only one steady state corresponding to the global minimum of the quadratic form $y^T W y - 2b^T y$ *over the set of* N*-dimensional binary vectors; (ii) algorithm (3) is stable; and (iii) the necessary number of steps needed to achieve the steady state (transient time) can be upperbounded by the following expression*

$$TR \leq \frac{N^2\|W\| + 2\sqrt{N^3}\|y\| + N\|W^{-1}\|\|b\|^2}{4k}, \tag{6}$$

where $\|\cdot\|$ *refers to the Euclidean norm.*

Here we only concentrate on demonstrating the fact that MVQF is Lyapunov function of the new algorithm which minimizes it, the detailed proof of Theorem 1, involving the globality of the solution can be found in [3].

To embark on the proof of minimization we need the following lemma.

LEMMA 1 *Let* $y(k+1) = \varphi(y(k))$ *a nonlinear recursion defined over the state space* $y \in Y$*. If there exists a function (the so-called Lyapunov function)* $L(y)$ *for which the following properties hold*

1. $L(y)$ *is bounded* $\exists A, B$: $A \leq L(y) \leq B \ \forall y \in Y$
2. $\Delta L(k) := L(y(k+1)) - L(y(k)) < 0 \ \forall y \in Y$ *then the recursion* $y(k+1) = \varphi(y(k))$ *converges to one of the local minima of* $L(y)$.

Based on this lemma the convergence properties of algorithm can be easily proven as follows: Analysing the change of the quadratic form due to the state transitions, we obtain expression

$$\Delta Q(k) := Q(k+1) - Q(k) = \Delta y_i^2(k) W_{ii} + 2\Delta y_i(k)\left\{\sum W_{ij} y_j(k) - b_i\right\},$$

where $Q(k) := y^T(k)Wy(k) - 2b^Ty(k)$ and $\Delta y_i(k) := y(k+1) - y(k)$. If there is a state transition then $y_i(k)$ can change from -1 to $+1$ or vice versa. Let us deal with the state transition from -1 to $+1$, which results in

$$\Delta Q(k) = 4W_{ii} + 4\left\{\sum_j W_{ij}y_j(k) - b_i\right\}.$$

Owing to the hysteresis type of nonlinearity (4) this state transition can only occur if $\sum_j W_{ij}y_j(k) - b_i \leq -r_i = (W_{ii} + k)$ which provides the following bound on $\Delta Q(k)$:

$$\Delta Q(k) \leq -4k < 0.$$

Now it is easy to verify that the same bound can be obtained for each i in the case of a $+1$ to -1 state transition, therefore the first condition for $Q(y)$ being a Lyapunov function of algorithm (4) is satisfied. $Q(y)$ can easily be lowerbounded by using the Schwarz inequality and taking into account that it has one global minimum over R^N in the point $m = W^{-1}B$ with value $Q(m) = m^TWm - 2b^Tm = -m^Tb - b^TW^{-1}b$. Therefore

$$Q(y) \geq -b^TW^{-1}b \geq -||b||^2||W^{-1}||\forall y \in \{-1,1\}^N$$

which implies the fulfillment of the second condition for $Q(y)$ being a Lyapunov function.

Taking into account that $y \in \{-1,1\}^N$, $Q(y)$ can be upperbounded as $Q(y) \leq ||y||^2||W|| + 2||b||||y|| = N||W|| + 2\sqrt{N}||b||$. Hence the total variation of Q is bounded

$$V_Q \leq N||W|| + 2\sqrt{N}||b|| + ||W^{-1}||||b||^2.$$

As a result, the necessary number of steps needed to achieve the global minimum (TR) can be upperbounded in the following fashion:

$$TR \leq N\frac{N||W|| + 2||b||\sqrt{N} + ||W^{-1}||||b||^2}{4k}.$$

The factor N reflects the fact that we are working in the sequential mode of operation, thus in the worst case it can take N steps until a component changes its value.

2.1. Application of the Modified Hopfield Net to the Detection Problem

In digital communication theory, detection of linearly distorted signals under Gaussian noise is of primary importance. Efficient detection algorithms make possible to implement low bit error rate (BER) receivers in QAM systems. Whereas traditional system design tried to keep BER at low level by using channel equalizers, the new quadratic minimizer as a detector can perform the optimal Bayesian detection rule.

The problem of optimal detection can be formulated as follows (see *Fig. 1*):

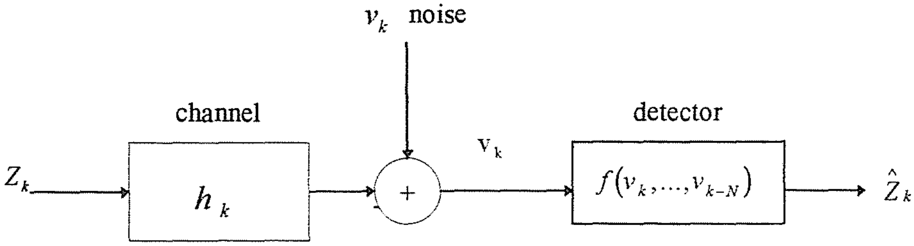

Fig. 1. Digital communication system with linearly distorted channel and Gaussian noise

where

Z_k	:	binary independent identically distributed random variables, where $k = 1, \dots, N$;
h_k	:	discrete impulse response of the channel, where $k = 0, \dots, M$ otherwise;
ν_k	:	Gaussian noise sequence $E\nu_k = 0$, $\forall k$, $E\nu_k\nu_l = K_{kl} = c_{\lvert k-l \rvert}$ and $E\nu_k^2 = N_0$;
$v_k = \sum_{n=0}^{M} h_n Z_{k-n} + \nu_k$	:	received sequence, where $k = 0, \dots, M + N$;
$\hat{Z}_k = f(v_k, \dots, v_{k-N})$	:	detected sequence;
N	:	the length of the transmitted sequence;
M	:	the length of the channel memory.

It is easy to see that the optimal detection reduces to the global minimization of a quadratic form given by the following expression.

$$\hat{y} : \max_y \frac{1}{\sqrt{2\pi \det K}} \exp\left(-\frac{1}{2}(v - Hy)^T K^{-1}(v - Hy)\right) =$$

$$\min_y (v - Hy)^T K^{-1}(v - Hy) = \min_y (y^T W y - 2b^T y) .$$

This prompts us to apply the modified Hopfield algorithm as an optimal detector given that the conditions listed in Theorem 1 are fulfilled. It can be easily proven [3] that these conditions are not restrictive at any rate, as far as a typical communication scenario is concerned ($h_0 > h_i$, $i = 1, \dots, M$).

The following figure shows some simulation results when the channel characteristics are $h_{-2} = 0.05$, $h_{-1} = 0.1$, $h_0 = 1$, $h_1 = 0.1$, $h_2 = 0.05$ and the noise is white Gaussian with $E\nu_l = 0$ and $\sigma^2 = 0.01$. The optimum

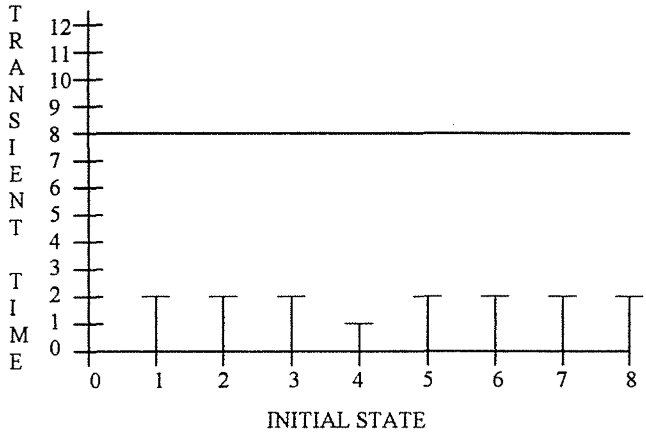

Fig. 2. Transient time of the detection by the new algorithm

detection rule was calculated for all possible three dimensional binary input vectors yielding the following convergence times.

As can be seen, this convergence time is far below the complexity given by the exhaustive search. Therefore, neural based optimization systems can be successfully applied in communication theory where the underlying optimization problems do not lend themselves to easy solutions by traditional methods.

3. Optimum Resource Management by Feedforward Neural Networks

As was detailed earlier, optimum resource management is concerned with evaluating the tail of the aggregated load [5,6] in the form of $P\left(\lim_{t\to\infty}\sum_{j=1}^{J} X_j(t) > C\right) < e^{-\gamma}$, where $X_j(t)$ represents the random load presented expressed in number of work unit/time unit by the sources of the system. C denotes the capacity of the system defined in terms of how many work units the system can handle during a time unit.

One can approach the problem by assuming memoryless independent sources [6,7]. In this case the formula above reduces to $P\left(\sum_{j=1}^{J} X_j > C\right) <$ $< e^{-\gamma}$ which allows the use of traditional statistical inequalities, such as the Chernoff and Hoeffding bounds. The Chernoff bound yields a relatively sharp estimation of the tail in the form of

$$P\left(\sum_{j=1}^{J} X_j > C\right) \le e^{\sum^{J} u_j(s^*)-s^*C},$$

where $\mu_j(s) = \ln Ee^{sX_j}$ and $s^* : \sum_{j=1}^{J} \frac{d\mu_j(s)}{ds} = C$.

Based on this bound CAC can be performed as $\sum_{j=1}^{J} \mu_j(s^*) - s^*C \leq$ $\leq -\gamma$. One can run into problem though by calculating the optimum s^*. This problem can be tiresome when the number of users are changing frequently, leading to numerous re-optimization of parameter s^*.

The Hoeffding inequality does not need the knowledge of the logarithmic moment generating function $\mu(s)$. The trade off for simplicity is the rough nature of this estimation given in the form of

$$P\left(\sum_{j=1}^{J} X_j > C - \sum_{j=1}^{J} m_j\right) \leq e^{-2\left(C - \sum_{j=1}^{J} m_j\right)^2 \Big/ \sum_{j=1}^{J}(b_j - a_j)^2},$$

where $a_j, b_j : P(a_j \leq X \leq b_j) = 1$. In spite of the fact that both upperbounds allow simple resource management, the system utilization may not be optimal due to approximate nature of the bounds. Therefore, other methods should be used for tail estimation.

Neural networks can be of help when tail estimation is reduced to a set separation problem. In this case the users are assumed to be On/Off type with Bernoulli distribution $(P(X_j = 0) = 1 - \frac{m_j}{h_j} P(X_j = h_j) = \frac{m_j}{h_j})$ and they are divided into classes $i = 1, \ldots, M$ with regard to their parameters m_i, h_i.

Users from the same class are supposed to be homogeneous having the same traffic characteristics. The system can be described by a traffic state vector $\overline{n} = (n_1, \ldots, n_i, \ldots, n_M)$ where the component n_i denotes the number of users being present from the ith class. Then CAC can be interpreted as a dichotomy in the traffic space expanded by vectors n (see *Fig. 3*)

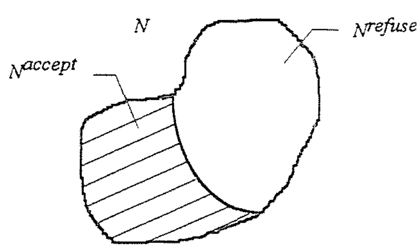

Fig. 3. The dichotomy of the traffic state space defined by CAC

which is generated by the following inequality .

$$P\left(\sum_{i=1}^{M} n_i X_i > C\right) \leq e^{-\gamma} . \tag{7}$$

As the calculation of (7) does not lend itself to numerical tractability, the task of CAC is to find a good approximation of the separation surface (see *Fig. 4*) allowing simple admission algorithm under the following constraints:

- $N^{\text{appr.accept}} \subseteq N^{\text{accept}}$
- the number of lost calls should be minimized (for a measure μ, $\min \mu(N^{\text{accept}} - N^{\text{appr.accept}})$).

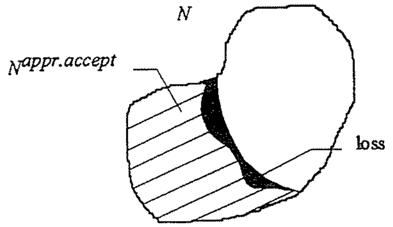

Fig. 4. Approximating the separation surface

An efficient approximation of the separation surface can be obtained by using a polygonal surface (see *Fig. 5*).

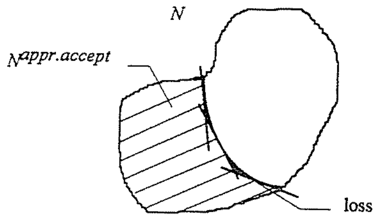

Fig. 5. Polygonal approximation of the separation surface

This polygonal approximation can be carried out by a two-layer neural network, in which the neurons in the first layer introduce separations by individual hyperplanes, whereas the single neuron in the second layer carries out an OR function to unite the individual separations, as indicated by *Fig. 6*:

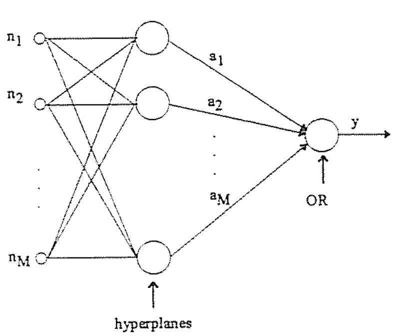

Fig. 6. A two-layer neural network carrying out the polygonal approximation

The input-output mapping of the corresponding neural network is given by (8)

$$y = \operatorname{sgn}\left\{\sum_{i=1}^{K} a_i\left(\sum_{j=1}^{M} b_{ij}x_j - b_{i0}\right) - a_0\right\} \tag{8}$$

and the weights can be optimized according to the following criterion:

$$w_{\text{opt}} : \min_{w} \mu(N^{\text{accept}} - N^{\text{appr.accept}}) ,$$

where μ is measure defined over the state space (e.g. $\mu(A)$ = the number of points which lie within A).

In the case of Markovian users, the following stochastic differential equation characterizes the system which now forms a queue

$$q(k+1) = [q(k)-1]^{+} + \sum_{j=1}^{J} X_j(k) .$$

The objective is to evaluate the tail of the stationary queue length distribution, which determines the cell loss probability.

$$\pi_i := \lim_{k\to\infty} P(q(k)=i) \quad P_{\text{cell loss}} = P_{\text{buffer overflow}} = \sum_{i\geq L} \pi_i .$$

An efficient estimate of the tail can be obtained by estimating the PERRON FROBENIOUS eigenvalue [5], yielding

$$\beta = 1 + 2(1-\rho)/\left\{\sum_{i}\left(\frac{Et_{\text{on}}^2}{Et_{\text{off}}^2} + \frac{Et_{\text{off}}^2}{Et_{\text{on}}^2} - 2\right)\frac{Et_{\text{on}}}{Et_{\text{on}}+Et_{\text{off}}}\left(1 - \frac{Et_{\text{on}}}{Et_{\text{on}}+Et_{\text{off}}}\right)^2 Et_{\text{on}} - \frac{Et_{\text{on}}}{Et_{\text{on}}+Et_{\text{off}}}\left(1-\frac{Et_{\text{on}}}{Et_{\text{on}}+Et_{\text{off}}}\right)\right\} .$$

CAC can then be performed based on the geometrical tail.

3.1. Simulation Results

Some numerical results are indicated in the next figure, where the admission region is shown in the case of heterogeneous traffic including two traffic classes.

As one can see, the neuron based management algorithms present the best approximation of the theoretically calculated separation surface achieving the highest system utilization.

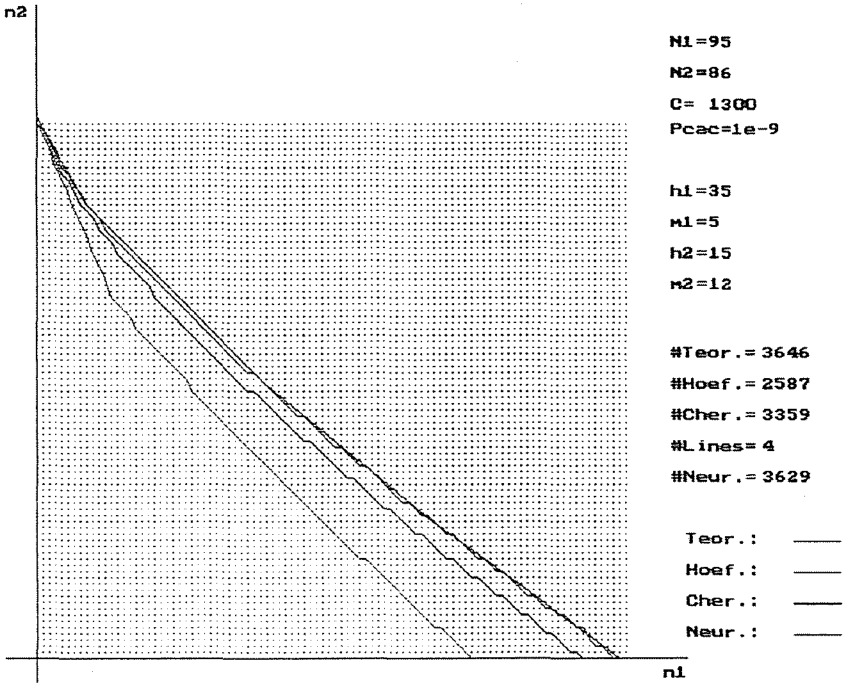

Fig. 7. Admission regions in the case of heterogeneous traffic obtained by neural set-separation, Hoeffding inequality and Chernoff inequality

4. Summary

The capability of solving hard optimization by neural networks was proven in two areas (i) the global minimization of discrete quadratic forms and (ii) in optimum resource management problems. In both cases, fast and low complexity solutions were achieved by neural architectures. The obtained modified Hopfield net and two-layer feedforward network can be successfully applied to the problem of detection and call admission control, respectively.

Acknowledgement

The research reported here was performed in the framework of the RTD USA BME Laboratories at the Department of Telecommunications. Part of the work was funded by project COP 579 sponsored by the European Union.

References

[1] HOPFIELD, J. J.: Neural Networks and Physical Systems with Emergent Collective Computational Abilities. *Proceedings of the National Academy of Sciences of the United States*, Vol. 79, 1982, pp. 2554–2558.

[2] HOPFIELD, J. J. – TANK, D. W.: Neural Computation of Decision in Optimization Problems, *Biological Cybernetics*, Vol. 52, 1985, pp. 141–152.

[3] LEVENDOVSZKY, J. – MOMMAERTS, W. – VAN DER MEULEN, E. C.: Hysteretic Neural Networks for Global Optimization of Quadratic Forms, *Neural Network World*, 1992, Vol. 2, No. 5, pp. 475–496.

[4] LEVENDOVSZKY, J. – MOMMAERTS, W. – VAN DER MEULEN, E. C.: Neural Networks with Hysteresis Type of Nonlinearity Exhibit Global Optimization Property, *Lecture Notes in Computer Science*, Vol. 540, *Artificial Neural Networks, International Workshop IWANN'91*, ed. by Prieto, A., Springer- Verlag, 1991.

[5] LEVENDOVSZKY, J. – IMRE, S. – VAN DER MEULEN, E C. – PAP, L. – VARGA, B.: Call Admission Control of ATM Networks Based on Modulated Markov Chains, *Journal on Communication dedicated to ATM Networks*, Vol. XLVII, pp. 19–24, March.

[6] LEVENDOVSZKY, J. – IMRE, S. – VAN DER MEULEN, E. C. – VARGA, B.: Call Admission Control for ATM Networks by Source Distribution Transformation, *Participants Proceedings of 4th IFIP Workshop on Performance Modelling of ATM Networks*, Ilkley, UK, July 9-12, 1996, pp. 17/1-17/12.

[7] LEVENDOVSZKY, J. – VAN DER MEULEN, E. C.: Tail Estimation by Statistical Bounds and Neural Networks, *Proceedings of 17th Benelux Symposium on Information Theory*, Enschede, The Netherlands, May 30-31, 1996, pp. 137–145.

[8] LEVENDOVSZKY, J. – IMRE, S. – VAN DER MEULEN, E. C. – PAP, L. – POZGAI, P. – VARGA, B.: Tail Distribution Estimation for Call Admission in ATM Networks, *Participants Proceedings of 3th IFIP Workshop and Performance Evaluation and Modelling of ATM Networks*, Ilkley, West Yorkshire, UK, July 1995.

[9] AARTS, E. – KORST, J.: Simulated Annealing and Boltzmann Machines. John Wiley and Sons, Chichester, 1989.

[10] LUCKY, R. W. – SALZ, J. – WELDON, E. J. Jr.: Principles of Data Communication, McGraw-Hill, New York. 1968.

[11] DUDA, R. O. – HART, P. E.: Pattern Classification and Scene Analysis. John Wiley & Sons, New York, 1973.

[12] PROAKIS, J. G.: Digital Communications. McGraw-Hill, Singapore, 1983.

[13] VITERBI, A. J. – OMURA, J. K.: Principle of Digital Communication and Coding. McGraw-Hill, New York, 1979.

[14] AMARI, S. – MAGINU, K.: Statistical Neurodynamics of Associative Memory, *Neural Networks*, Vol. 1, 1989, pp. 63–73.