A CONNECTIONIST APPROACH IN BAYESIAN CLASSIFICATION

Luminita State
Department of Computer Science, University of Pitesti, Pitesti, Romania

Catalina Cocianu
Department of Computer Science, Academy of Economic Studies, Bucharest, Romania

Panayiotis Vlamos
Department of Computer Science, Ionian University, Corfu, Greece

Viorica Stefanescu
Department of Mathematics, Academy of Economic Studies, Bucharest, Romania

Keywords: Hidden Markov Models, learning by examples, Bayesian classification, training algorithm, neural computation.

Abstract: The research reported in the paper aims the development of a suitable neural architecture for implementing the Bayesian procedure in solving pattern recognition problems. The proposed neural system is based on an inhibitive competition installed among the hidden neurons of the computation layer. The local memories of the hidden neurons are computed adaptively according to an estimation model of the parameters of the Bayesian classifier. Also, the paper reports a series of qualitative attempts in analyzing the behavior of a new learning procedure of the parameters an HMM by modeling different types of stochastic dependencies on the space of states corresponding to the underlying finite automaton. The approach aims the development of some new methods in processing image and speech signals in solving pattern recognition problems. Basically, the attempts are stated in terms of weighting processes and deterministic/non deterministic Bayesian procedures.

1 PRELIMINARIES

Stochastic models represent a very promising approach to temporal pattern recognition. An important class of the stochastic models is based on Markovian state transition, two of the typical examples being the Markov model (MM) and the Hidden Markov Model (HMM). In a Markov model, the transition between states is governed by the transition probabilities, that is, the state sequence is a Markov process and the observable state is then directly observed as the output feature. However, usually, there are two sorts of variable to be taken into consideration, namely the *manifest* variables which can be directly observed and *latent* variables that are hidden to the observer. The HMM model is based on a doubly stochastic process, one producing an (unobservable) state and another producing an observable feature sequence.

The doubly stochastic process is useful in coping with unpredictable variation of the observed patterns and its design requires a learning phase when the parameters of both, the state transition and emission distributions have to be estimated from the observed data. The trained HMM can be then used for the retrieving (recognition) phase when the test sequence (complete or incomplete) observations have to be recognized.

The latent structure of observable phenomenon is modeled in terms of a finite automaton Q, the observable variable being thought as the output

State L., Cocianu C., Vlamos P. and Stefanescu V. (2007).
A CONNECTIONIST APPROACH IN BAYESIAN CLASSIFICATION.
In *Proceedings of the Ninth International Conference on Enterprise Information Systems - AIDSS*, pages 185-190
DOI: 10.5220/0002346401850190

produced by the states of Q. Both evolutions, in the spaces of non observable as well as in the space of observable variables, are assumed to be governed by probabilistic laws.

In the sequel, we denote by $(\Lambda_n)_{n\geq 0}$ the stochastic process describing the hidden evolution and by $(X_n)_{n\geq 0}$ the stochastic process corresponding to the observable evolution.

Let Q be the set of states of the underlying finite automaton; $|Q| = m$. We denote by τ_n the probability distribution on Q at the moment n. Let $(\Omega, \Im, P)$ be a probability space, $(\aleph, C, \sigma)$ be a measure space, where σ is a σ-finite measure.

We assume that $\rho : Q \to C^*$ is a σ-experiment, that is for any $q \in Q$, $\rho(q) << \sigma$, where C^* is the set of all probability measures defined on the σ algebra C. Let $f_q(.)$ be a measurable version of the Radon-Nycodim derivative $f_q = \frac{d\rho(q)}{d\sigma}$. The output of each state $q \in Q$ is represented by the random element $X : \Omega \to \aleph$ of density function $f_q(.)$. Let ξ be the *apriori* probability distribution on Q; for any $q \in Q$, $\xi(q)$ is the subjective credibility that q is the true emitting state at any moment. We assume that $\forall q \in Q, \xi(q) \neq 0$. The conclusions on the hidden evolution are derived using the Bayesian procedure when the *apriori* probability distribution ξ and the set of density functions $(f_{n,q}, q \in Q)$ are known.

If $L : Q \times Q \to [0,\infty)$ is a risk function, then, for any $q, q^* \in Q$, $L(q, q^*)$ represents the cost implied by taking the output emitted by q as being emitted by q^*. The outputs of the automaton are represented by the sequence of random elements $(X_n)_{n\geq 0}$, where the output at the moment n, X_n is distributed $\rho(q_n)$ if it was emitted by the state q_n.

A random decision procedure is an element of

$$R = \left\{ t \,/\, t \in \left([0,1]^Q\right)^{\aleph} \right\},$$

where, for any $t \in R, q \in Q, x \in \aleph$, $t(x)(q)$ is the probability of deciding that the output x is produced by the state q.

For any $t \in R$ we denote the expected risk by,

$$R(\xi, t, f) = \sum_{q\in Q} \sum_{\bar{q}\in Q} \int_{\aleph} \xi(q) L(q, \bar{q}) t_{\bar{q}}(x) f_q(x) \sigma(dx)$$

The Bayesian decision procedure $\tilde{t} \in R$ assures the minimum risk that is, $R(\xi, \tilde{t}, f) = \inf_{t\in R} R(\xi, t, f) \underline{\underline{\Delta}} \Phi(\xi, f)$ and it is given by,

$$(1)\ \tilde{t}_{\bar{q}}(x) = \begin{cases} 1, & T(\bar{q}, x) < \min\limits_{q^* \in Q\setminus\{\bar{q}\}} T(q^*, x) \\ 0, & T(\bar{q}, x) > \min\limits_{q^* \in Q\setminus\{\bar{q}\}} T(q^*, x) \\ \alpha_{\bar{q}}, & T(\bar{q}, x) = \min\limits_{q^* \in Q\setminus\{\bar{q}\}} T(q^*, x) \end{cases}, \text{ where}$$

$$(2)\ T(\bar{q}, x) = \sum_{q\in Q} \xi(q) L(q, \bar{q}) f_q(x), \ \sum_{\bar{q}\in A} \alpha_{\bar{q}} = 1,$$

$$\forall \bar{q} \in A, \alpha_{\bar{q}} \geq 0$$

$$\text{and } A = \left\{ \bar{q} \,/\, \bar{q} \in Q, T(\bar{q}, x) = \min_{q^* \in Q\setminus\{\bar{q}\}} T(q^*, x) \right\}.$$

The true evolution in the space Q of non observable variables is governed by probabilistic laws, $(\tau_n)_{n\geq 0}$, where τ_n represents the probability distribution on Q at the moment n.

Let $(u_n)_{n\geq 0}$ be a sequence of subjective utilities assigned to the states of the automaton; $\forall n \geq 0, u_n : Q \to [0,\infty)$. We assume that, for any $n \geq 1, \sum_{q\in Q} u_n(q) \neq 0$. For any $n \geq 0$ and $q \in Q$, $u_n(q)$ stands for the subjective utility assigned to the state q at the moment n. Typically, $u_n(q)$ can be taken as the relative emitting frequency of the state q during the time interval $[0, n]$.

In case the HMM evolution is directly observable of a certain time interval $[1, N]$, that is a sequence of N-realizations of both processes $(\Lambda_n)_{n\geq 0}$ and $(X_n)_{n\geq 0}$ are available to the experimenter, we get a learning sequence of length N which can be used to estimate the hidden evolution on Q as well as to derive estimations for the conditional density functions $(f_{n,q}, q \in Q)$. Let $(g_n)_{n\geq 1}$ be a sequence of measurable functions, $g_n : \aleph \times \aleph \to [0,\infty)$, $\forall n \geq 1$, such that the following regularity conditions hold,

$$(A_1) \quad \forall n \geq 1, \int_{\aleph} g_n(x, y) \sigma(dy) = 1 \ , \ a.s. - \sigma$$

$$(A_2) \quad \forall n \geq 1, \forall x, y \in \aleph, 0 \leq g_n(x, y) \leq 1$$

$(A_3)\quad \forall q \in Q,\ \lim_{n\to\infty} E_q\left(g_n(x,X)\right)=$
$\lim_{n\to\infty}\int_{\aleph} g_n(x,y) f_q(y)\sigma(dy) = f_q(x),\ a.s.-\sigma$

Our method is a supervised technique based on the learning sequence $S=\left((\Lambda_n,X_n)/\, n\geq 1\right)$, where the true probability distribution τ_n is approximated by a weighting process $\left(\xi_n(q), q\in Q\right)_{n\geq 0}$ defined by $\xi_n(q)=\frac{\xi(q)u_n(q)}{\sum_{\bar q\in Q}\xi(\bar q)u_n(\bar q)}$ representing the guess that q is the emitting state at the moment n. The decision procedure $\tilde t_n^*$ is defined by (1) in terms of $\xi_n(q)$ and $f_{n,q}(x)=\frac{1}{n\xi_n(q)}\sum_{j=1}^{n}\delta(\Lambda_j,q)g_n(x,X_j)$, where $\delta(q,\bar q)=\begin{cases}1, & q=\bar q\\ 0, & q\neq \bar q\end{cases}$. The criterion function $T(\bar q,x)$ given by (2) is replaced by $T(\bar q,x)=\sum_{q\in Q}\xi_n(q)L(q,\bar q)f_{n,q}(x)$.

2 QUALITATIVE ANALYSIS OF THE LEARNING SCHEME

Let $R_\xi\left(\tilde t_n^*\right)=E\left(R\left(\xi,\tilde t_n^*,f\right)\right)$ be the expected risk corresponding to the random decision procedure $\tilde t_n^*$ when ξ is the true probability distribution on Q and $f=\left(f_{n,q}, q\in Q\right)$ is the set of output density functions.

Theorem 1. Let $(g_n)_{n\geq 0}$ be a sequence of measurable functions such that the assumptions A_1, A_2, A_3, A_4 hold, where,
(A_4) for any $k\geq 1$, $q\in Q, x\in\aleph$,

$$E_q\left(g_k(x,X)\right)=f_q(x).$$

If $S=\left((\Lambda_n,X_n)/\, n\geq 1\right)$ is a learning sequence such that the random elements $(\Lambda_n,X_n), n\geq 1$ are independent, Λ_n is distributed ξ and X_n is distributed f_q if $\Lambda_n=q$, then,

$$\lim_{n\to\infty} R_\xi\left(\tilde t_n^*\right)=\Phi(\xi,f).$$

Proof: The conclusion can be established using straightforward computations and invoking the strong law of large numbers and the dominated convergence theorem.

Theorem 2. Let $S=\left((\Lambda_n,X_n)/\, n\geq 1\right)$ be a learning sequence such that the random elements $(\Lambda_n,X_n), n\geq 1$ are independent, Λ_n is distributed τ_n and X_n is distributed f_q if $\Lambda_n=q$. If for the sequence $(g_n)_{n\geq 0}$, the assumptions A_1, A_2, A_3, A_4 hold and, for any $q\in Q$, $\lim_{n\to\infty}\frac{1}{n}\sum_{j=1}^{n}\tau_j(q)=\tau(q)$, then,

$$\lim_{n\to\infty} E\left(R\left(\tau,\tilde t_n^*,f\right)\right)=\Phi(\xi,f).$$

Proof: The following series of equations can be derived,

$$0\leq E\left(R\left(\tau,\tilde t_n^*,f\right)\right)-\Phi(\tau,f)\leq$$
$$\leq \sum_{q\in Q}\bar L_q\int_{\aleph} E\left|\xi_n(q)f_{n,q}(x)-\tau(q)f_q(x)\right|\sigma(dx)+$$
$$+\sum_{q\in Q}\sum_{\bar q\in Q}L(q,\bar q)\left[\frac{1}{n}\sum_{j=1}^{n}\tau_j(q)-\tau(q)\right]$$
$$\int_{\aleph} t_{\bar q}(\tau,f,x)f_q(x)\sigma(dx)$$

Obviously, the second term converges to 0 when $n\to\infty$.

Also, using the strong law of large numbers, we obtain,

$$\lim_{n\to\infty}\frac{1}{n}\sum_{j=1}^{n}\left[\delta(\Lambda_j,q)g_n(x,X_j)-\tau_j(q)f_q(x)\right]=0\ \text{a.s.-}P$$

for any $x\in\aleph, q\in Q$.

Using the dominated convergence theorem, we get

$$\lim_{n\to\infty}E\left\{\left|\xi_n(q)f_{n,q}(x)-\left[\frac{1}{n}\sum_{j=1}^{n}\tau_j(q)\right]f_q(x)\right|\right\}=0\ \text{a.s.-}P$$

for any $x\in\aleph, q\in Q$ which finally implies that, for any $q\in Q$,

$$\lim_{n\to\infty} \int_{x\in\aleph} E\left\{\left|\xi_n(q) f_{n,q}(x) - \left[\frac{1}{n}\sum_{j=1}^{n}\tau_j(q)\right] f_q(x)\right|\right\}\sigma(dx) = 0$$

which implies the conclusion of Theorem 2.

Theorem 3. Assume that the conditions mentioned in theorem 2 hold. If, for any $q \in Q$, $\lim\limits_{n\to\infty} \tau_n(q) = \tau(q)$, then,

$$\lim_{n\to\infty} E\left(R\left(\tau_n, \tilde{t}_n^*, f\right)\right) = \Phi(\tau, f) .$$

Proof: Since

$$E\left(R\left(\tau_n, \tilde{t}_n^*, f\right)\right) - \Phi(\tau, f) = E\left[R\left(\tau_n, \tilde{t}_n^*, f\right) - \Phi(\tau_n, f)\right] + \\ + \left[\Phi(\tau_n, f) - \Phi(\tau, f)\right]$$

we obtain,

$$\left|\Phi(\tau_n, f) - \Phi(\tau, f)\right| \le \\ \le \sum_{q\in Q} \overline{L}_q \int_{\aleph} f_q(x) \sum_{\overline{q}\in Q}\left|\tilde{t}_{n,\overline{q}}^*(x) - t_{\overline{q}}(\tau, f, x)\right|\sigma(dx) + \\ + \sum_{q\in Q}\overline{L}_q\left|\tau_n(q) - \tau(q)\right| .$$

Using the dominated convergence theorem, we get

$$\lim_{n\to\infty} \sum_{q\in Q}\overline{L}_q \int_{\aleph} f_q(x) \sum_{\overline{q}\in Q}\left|\tilde{t}_{n,\overline{q}}^*(x) - t_{\overline{q}}(\tau, f, x)\right|\sigma(dx) = 0$$

and, consequently,

$$\lim_{n\to\infty}\left[\Phi(\tau_n, f) - \Phi(\tau, f)\right] = 0 .$$

Using Theorem 2, the definition of procedures $t(\tau_n, f), t(\tau, f)$ and dominated convergence theorem, we get,

$$\lim_{n\to\infty}\left[\Phi(\tau, f) - R\left(\tau, t(\tau_n, f), f\right)\right] = 0$$

and

$$\lim_{n\to\infty} E\left\{\sum_{q\in Q}\sum_{\overline{q}\in Q} L(q,\overline{q})\left[\int_{\aleph}\tau(q) f_q(x)\tilde{t}_{n,\overline{q}}^*(x)\sigma(dx) - \right.\right. \\ \left.\left. - \int_{\aleph}\tau(q) f_q(x) t_{\overline{q}}(\tau_n, f, x)\sigma(dx)\right]\right\} = 0$$

Since

$$\left\langle E\left\{\sum_{q\in Q}\sum_{\overline{q}\in Q}\left[L(q,\overline{q})\left(\tau(q) - \tau_n(q)\right)\int_{\aleph} f_q(x) t_{\overline{q}}(\tau_n, f, x)\sigma(dx)\right]\right\} = 0\right\rangle$$ and

$$\left|E\left\{\sum_{q\in Q}\sum_{\overline{q}\in Q} L(q,\overline{q})\left[\int_{\aleph}\left(\tau_n(q) f_q(x) - \xi_n(q) f_{n,q}(x)\right)\tilde{t}_{n,\overline{q}}^*(x)\sigma(dx) + \right.\right.\right. \\ \left.\left.\left. \int_{\aleph}\left(\xi_n(q) f_{n,q}(x) - \tau(q) f_q(x)\right)\tilde{t}_{n,\overline{q}}^*(x)\sigma(dx)\right]\right\}\right| \le \sum_{q\in Q}\overline{L}_q\left|\tau_n(q) - \tau(q)\right|$$

we finally get,

$$\lim_{n\to\infty} E\left\{\sum_{q\in Q}\sum_{\overline{q}\in Q} L(q,\overline{q})\left[\int_{\aleph}\left(\tau_n(q) f_q(x) - \xi_n(q) f_{n,q}(x)\right)\tilde{t}_{n,\overline{q}}^*(x)\sigma(dx) + \right.\right. \\ \left.\left. \int_{\aleph}\left(\xi_n(q) f_{n,q}(x) - \tau(q) f_q(x)\right)\tilde{t}_{n,\overline{q}}^*(x)\sigma(dx)\right]\right\} = 0 ,$$

which implies

$$\lim_{n\to\infty} E\left[R\left(\tau_n, \tilde{t}_n^*, f\right) - \Phi(\tau, f)\right] = 0 .$$

Let us assume that $\aleph$ is a denumerable set, $\sigma(x) = 1, \forall x \in \aleph$. Obviously, taking $(g_n)_{n\ge 0}$ such that for any $n \ge 0$ and for any $x, y \in \aleph$, $g_n(x, y) = \delta(x, y)$, the conditions A_1, A_2, A_3 hold. Since for any $q \in Q$, $k \ge 1$, $E_q\left(g_k(x, X)\right) = \sum_{y\in\aleph} g_k(x, y) f_q(y) = f_q(x)$, we get that A_4 also holds.

Theorem 4. Let $S = \left((\Lambda_n, X_n)/\, n \ge 1\right)$ be a learning sequence such that $(\Lambda_n, n \ge 1)$ is a Markov chain of stationary transition probabilities having an unique recurrent class Q'. If $(X_n), n \ge 1$ are independent and X_n is distributed f_q if $\Lambda_n = q$, then

$$\lim_{n\to\infty} E\left(R\left(\tau, \tilde{t}_n^*, f\right)\right) = \Phi(\tau, f) ,$$

where τ is the probability distributions of Λ_1 .

Proof: For $q \in Q$, $x \in \aleph$, we define,

$$(3)\ f\left((\Lambda, x)\right) = \delta(\Lambda, q)\delta(X, x) .$$

Obviously, f is $\Im_{(\Lambda_1, X_1)}$-measurable. Since

$$E\left\{f(\Lambda_1, X_1)\right\} = \\ = \sum_{\overline{x}\in\aleph}\sum_{\overline{q}\in Q}\delta(q,\overline{q})\delta(x,\overline{x}) f_{\overline{q}}(\overline{x})\tau(\overline{q}) = \tau(q) f_q(x) < \infty ,$$

we obtain,

$$(4)\ E\left\{f(\Lambda_1, X_1)\right\} = \tau(q) f_q(x) .$$

Also, the series $\sum_{(\bar{q},\bar{x})\in Q\times\aleph} r_{(q^*,x^*)}(\bar{q},\bar{x})$ converge uniformly in (q^*,x^*). We get that, for f defined by (3), the conditions of Theorem 1 hold.

Using (4), Theorem 1 and the dominated convergence theorem, we get

$$\lim_{n\to\infty}\int_{\aleph} E\left\|\xi_n(q)f_{n,q}(x)-\tau(q)f_q(x)\right\|\sigma(dx)=0,$$

which implies

$$\lim_{n\to\infty}\sum_{q\in Q}\bar{L}_q\int_{\aleph} E\left\|\xi_n(q)f_{n,q}(x)-\tau(q)f_q(x)\right\|\sigma(dx)=0.$$

Finally, since

$$0\le E\left[R\left(\tau,\tilde{t}_n^*,f\right)\right]-\Phi(\tau,f)\le$$
$$\le\lim_{n\to\infty}\sum_{q\in Q}\bar{L}_q\int_{\aleph} E\left[\left|\xi_n(q)f_{n,q}(x)-\tau(q)f_q(x)\right|\right]\sigma(dx)$$

we get

$$\lim_{n\to\infty}E\left(R\left(\tau,\tilde{t}_n^*,f\right)\right)=\Phi(\tau,f).$$

3 NEURAL IMPLEMENTATION

We assume that $\aleph=R^d$. Then the neural architecture consists of the layers F_X, F_H of d and respectively $|Q|$ neurons. The neurons of the input layer F_X have no local memory, they distribute the corresponding inputs toward the neurons of the hidden layer F_H. Each neuron of F_H is assigned to one of the pattern classes from Q. For simplicity sake, we'll refere to each neuron of F_H by its corresponding pattern class.

The local memory of each neuron $q\in F_H$ consists of $\xi_n(q)$ and the parameters needed to compute $f_{n,q}$. The activation function of the neuron $q\in F_H$ at the moment n is $h_{n,q}(\mathbf{x})=f_{n,q}(\mathbf{x})\xi_n(q)$. The layer F_H is fully connected, the connection from q to $\bar{q}$ is weighted by $(-L(q,\bar{q}))$. Consequently, the input $\mathbf{x}=(x_1,...,x_d)$ applied to F_X induces the neural activations,

$$net(\bar{q},0)=-\sum_{q\in Q}\xi_n(q)L(q,\bar{q})f_{n,q}(\mathbf{x})=$$
$$=-T(\bar{q},\mathbf{x}),\ \bar{q}\in F_H.$$

The recognition task corresponds to the identification of the states $\bar{q}$ for which $T(\bar{q},\mathbf{x})$ is minimum. This task is solved by installing a discrete time competitive process among the neurons of F_H. Let $S_q(t)=f(net(q,t))$ be the output of the neuron $q\in F_H$ at the moment t, where the competition process starts at the moment 0 and the activation function f is given by $f(u)=\begin{cases}0, & u\ge 0\\ u, & u<0\end{cases}$. We denote by $S(t)=(S_q(t),q\in F_H)$ the state at the moment t. The initial state is $S(0)=(f(net(q,0)),q\in F_H)$.

The synaptic weights of the connections during the competition are,

$$w_{q,\bar{q}}=\begin{cases}1, & q=\bar{q}\\ -\varepsilon, & q\ne\bar{q}\end{cases},$$

where $\varepsilon>0$ is a vigilance parameter.

The update of the state is performed synchronously, that is, for any $q\in F_H$,

$$net(q,t+1)=S_q(t)-\varepsilon\sum_{\bar{q}\ne q}S_{\bar{q}}(t)=$$
$$=(1+\varepsilon)S_q(t)-\varepsilon\sum_{\bar{q}\in F_H}S_{\bar{q}}(t)$$
$$S_q(t+1)=f(net(q,t+1)).$$

The conclusions concerning the behavior of the competition in the space of states stem from the following arguments. Note that $S_q(t)\le 0$, for any $t\ge 0$ and $q\in F_H$.

1. If $S_q(t)=0$, then $h_q(t+1)\ge 0$, hence $S_q(t+1)=0$. Moreover, for any $t'\ge t$, $S_q(t')=0$.

2. Assume that for some $q\in F_H$, $t\ge 0$, $S_q(t)<0$. If $net(q,t+1)<0$ then

$$0\ge S_q(t+1)=S_q(t)-\varepsilon\sum_{q'\ne q}S_{q'}(t)\ge S_q(t)\text{ and}$$
$$S_q(t+1)=S_q(t)\text{ if and only if }S_{q'}(t)=0\text{ for all }q'\ne q.$$

3. Assume that for some $q,q'\in F_H$, $t\ge 0$, $S_{q'}(t)=S_q(t)<0$. Then

$$net(q,t+1)=(1+\varepsilon)S_q(t)-\varepsilon\sum_{\bar{q}\in F_H}S_{\bar{q}}(t),$$
$$net(q',t+1)=(1+\varepsilon)S_{q'}(t)-\varepsilon\sum_{\bar{q}\in F_H}S_{\bar{q}}(t),$$

that is $S_{q'}(t+1) = S_q(t+1) \le 0$ and, for any $t' \ge t$, $S_{q'}(t') = S_q(t') \le 0$.

4. Assume that for some $q, q' \in F_H$, $t \ge 0$, $S_{q'}(t) < S_q(t) < 0$. Then,

$$net(q, t+1) = (1+\varepsilon)S_q(t) - \varepsilon \sum_{\bar{q} \in F_H} S_{\bar{q}}(t),$$

$$net(q', t+1) = (1+\varepsilon)S_{q'}(t) - \varepsilon \sum_{\bar{q} \in F_H} S_{\bar{q}}(t),$$

that is $net(q, t+1) > net(q', t+1)$.

Obviously, if $net(q', t+1) \ge 0$ then $S_{q'}(t+1) = S_q(t+1) = 0$. Also, if $net(q, t+1) > 0 \ge net(q', t+1)$ then

$$S_{q'}(t+1) = net(q', t+1) \le 0 = S_q(t+1),$$

so we get $S_{q'}(t+1) \le S_q(t+1)$.

Finally, if $net(q, t+1) < 0$ then

$$net(q, t+1) = (1+\varepsilon)S_q(t) - \varepsilon \sum_{\bar{q} \in F_H} S_{\bar{q}}(t) >$$
$$> (1+\varepsilon)S_{q'}(t) - \varepsilon \sum_{\bar{q} \in F_H} S_{\bar{q}}(t) = net(q', t+1),$$

that is $S_{q'}(t+1) < S_q(t+1)$.

Consequently, if $S_{q'}(t) < S_q(t)$, then $S_{q'}(t+1) \le S_q(t+1)$. Moreover, for any $t' \ge t$, $S_{q'}(t') \le S_q(t')$.

From

$$S_q(t+1) - S_{q'}(t+1) = (1+\varepsilon)(S_q(t) - S_{q'}(t))$$

we get

$$S_q(t) - S_{q'}(t) = (1+\varepsilon)^t (S_q(0) - S_{q'}(0)),$$

that is if both components of the state vector were different from 0, for any $t \ge 0$, then $\lim_{t\to\infty}(S_q(t) - S_{q'}(t)) = \infty$, hence $\lim_{t\to\infty}(S_{q'}(t)) = -\infty$, which obviously contradicts the conclusion established by 2.

We arrived at the conclusion that there exists $t(q') \ge 0$ such that $S_{q'}(t) = 0$ for any $t \ge t(q')$.

5. Assume that for $q, q' \in F_H$, $0 < T(q, \mathbf{x}) < T(q', \mathbf{x})$, $S_{q'}(0) < S_q(0) < 0$.

Using the previously obtained conclusions, we get that, for any $t \ge 0$, $S_{q'}(t) \le S_q(t) \le 0$ and there exists $t(q') \ge 0$ such that $S_{q'}(t) = 0$ for any $t \ge t(q')$. Therefore, the competition installed by the above mentioned process among the neurons of F_H determines that the outputs of all neurons q' that received values $T(q', \mathbf{x}) > \min_{q \in F_H} T(q, \mathbf{x})$ are inhibited in a finite number of stages, that is there exists t_{fin} such that $S_{q'}(t_{fin}) \ne 0$ if and only if $T(q', \mathbf{x}) = \min_{q \in F_H} T(q, \mathbf{x})$.

Moreover, using the remark 3, we get that, for any $q', q'' \in F_H$ such that

$$T(q', \mathbf{x}) = T(q'', \mathbf{x}) = \min_{q \in F_H} T(q, \mathbf{x}),$$
$$S_{q'}(t_{fin}) = S_{q''}(t_{fin}) \ne 0$$

and for any $t \ge 0$, $S_{q'}(t) = S_{q''}(t)$.

The local memories of the hidden neurons are determined in a supervised way by adaptive learning algorithms using a learning sequence $S = ((\Lambda_n, X_n) / n \ge 1)$. The recurrent relations for $f_{n,q}, \xi_n(q), n \ge 1, q \in F_H$ are derived in terms of the particular relationships of $(u_n(q)), (g_n(\mathbf{x}, \mathbf{y}))$.

4 CONCLUSIONS

The supervised estimation techniques of the Bayesian decision procedure in pattern recognition presented in the paper were tested against data in automated speech recognition.

REFERENCES

Bishop, C., 1996. *Neural Networks for Pattern Recognition*, Oxford University Press

Devroye, L., Gyorfi, L., Lugosi, G., 1996. *A Probabilistic Theory of Pattern Recognition*, Springer Verlag

Husmeier, D., 2000 Learning Non-stationary Conditional Probability Distributions. In *Neural Networks, Vol. 13*

Lampinen, J., Vehtari,2001. A. Bayesian Approach for Neural Networks: Review and Case Studies. In *Neural Networks*, Vol. 14

State, L.. C. Cocianu, 2001. Information Based Algorithms in Signal Processing. In *Proceedings of SYNASC'2001*, Timisoara, 3-5 octombrie.