Normal Point Density Input Influence over a Finite One-Dimensional SOM

Ivo R. Draganov[1]

***Abstract* – In this paper an analysis is presented concerning the asymptotic state of the one-dimensional self-organizing map (SOM) with finite grid in the case of normal point distribution input. The SOM distortion measure is analyzed with its optimum found approximately. The results obtained are considered useful enough in wide variety of practical cases where fine tuning of the SOM is needed.**

***Keywords* – Self-Organizing Map, Normal Point Distribution, Distortion Measure.**

I. INTRODUCTION

It is well known fact that the area allocated for storing the most important feature set inside a self-organizing map (SOM) is proportional to the frequency of occurrence of that very same feature in the observations [1]. As the SOM structure tends to become very complex in the most of its real case applications often the magnification factor is used to describe heaping of feature vectors. It is simply the inverse of the point density around each neuron representing a cluster.

So far an investigation of the point density for the linear map is led in the presence of a very large number of codebook vectors over a finite area [2], [3]. It is revealed that the asymptotic point density is proportional to the probability of a certain feature vector occurring raised to some exponent depending of the number of neighbors including the winning neuron and some scalar factor.

In any case the initial neighbor function width may vary largely during the training process starting with huge values and ending with zero-order topology case – no neighbors except the winner are present. This boundary case is undesired since the learning process no longer maintains the order of the codebook vectors. The approximation accuracy of the probability of occurring for a feature and the minimum stability of ordering demanding more neighbor interactions are the two aspects to be balanced.

If we have no neighbors around the winner a simple scalar quantization case occurs. Then the power of the asymptotic function for the point density decreases, according to [3] bellow 1/3. Getting this power to higher values incrementally by trial and error approach seems a good solution but the following tendencies should be considered. If we try to approach the Bayesian classifier, i.e. to find the optimal classification border and the density functions of adjacent clusters are close to each other the latter could be replaced with any other pair of monotonic functions of densities. In such a case the practical SOM application is adaptable to simplification. The other important property is that when feature dimensionality is increased in the order of hundreds of components per vector the power is close to 1 [3]. Similar research on the change of this power is done in [4] when the neighbor function is Gaussian kernel and its normalized second moment is independent variable. The resulting range for the power value in this case is from 1/3 to 2/3. Analogous results are presented in [5].

What has not been investigated so far is the influence of the normal point density of the input over the asymptotic state of a finite one-dimensional SOM and its distortion measure. In [6] a typical practical challenge is given which can be solved by the approach presented here. In part 2 such analysis is presented and in part 3 some computational results are given. In part 4 a conclusion is made.

[1]Ivo R. Draganov is with the Faculty of Telecommunications, 8 Kliment Ohridski Blvd., 1000 Sofia, Bulgaria, E-mail: idraganov@tu-sofia.bg

II. SOM ANALYSIS WITH NORMAL POINT DENSITY INPUT

A. Asymptotic State of the One-Dimensional Finite-Grid SOM

Let one-dimensional feature space of x is considered. For our analysis to be correct the following assumptions should be granted: the number of points (feature vectors) must be large enough (e.g. by criteria given in [1]) and they must be stochastic variables so their differential probability for each cluster they fall into, i.e. the probability density $p(x)$ could be defined. The codebook vectors m_i usually form regular optimal configuration and thus can not be stochastic. Their number is typically low in any cluster as well.

Let suppose m_i and m_{i+1} are two neighboring points. A way of defining the point density is as $(m_{i+1} - m_i)^{-1}$ but it does not cover the samples around the boundaries of the clusters for which this density does not have meaning. So a better way of defining it is as the inverse of the width of the Voronoi set $[(m_{i+1} - m_i)/2]^{-1}$.

The input consists of samples $x(t) \in \Re, t = 0,1,2,...$ while the codebook is represented by $m_i(t) \in \Re, t = 0,1,2,..., i = 1,..., k.$ It is assumed $0 \le x(t) \le 1$. The one-dimensional SOM algorithm with at least one neighbor at each side according to [1] is given by:

$$m_i(t+1) = m_i(t) + \varepsilon(t)[x(t) - m_i(t)], \quad \text{for i} \in \mathrm{N_c},$$

$$m_i(t+1) = m_i(t) \quad \text{for} \quad \mathrm{i} \notin \mathrm{N_c},$$
$$c = \arg\min_{\mathrm{i}} \{|x(t) - m_i(t)|\}, \tag{1}$$
$$N_c = \{\max(l, c-1), c, \min(k, c+1)\},$$

where N_c is the neighbor set around node c and $\varepsilon(l)$ is the learning-rate factor. The Voronoi set V_i around m_i is defined as:

$$V_i = \left[\frac{m_{i-1}+m_i}{2}, \frac{m_i+m_{i+1}}{2}\right], V_1 = \left[0, \frac{m_1+m_2}{2}\right],$$
$$\mathrm{V_k} = \left[\frac{m_{k-1}+m_k}{2}, 1\right], \quad \text{for} \quad 1 < \mathrm{i} < \mathrm{k}, \tag{2}$$
$$U_i = V_{i-1} \cup V_i \cup V_{i+1}, U_1 = V_1 \cup V_2,$$
$$\mathrm{U_k} = V_{k-1} \cup V_k, \quad \text{for} \quad 1 < \mathrm{i} < \mathrm{k}.$$

In this case U_i is the set of such $x(t)$ which provoke changes in $m_i(t)$ during one learning step. Following (1) and (2) we get to the well known stationary equilibrium for m_i coinciding for the general case [1]:

$$m_i = E\{x \mid x \in U_i\}, \forall \mathrm{i}. \tag{3}$$

In other words every m_i becomes centroid of the probability mass for each U_i and then for 2 < i < (k-1) the limits for U_i are:

$$A_i = \frac{1}{2}(m_{i-2} + m_{i-1}),$$
$$B_i = \frac{1}{2}(m_{i+1} + m_{i+2}). \tag{4}$$

For i =1 and i = 2, A_i = 0, and for $i = k - 1$ and $i = k$, = 1.

The case investigated here concerns input data with the following distribution:

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-x_0)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}}, \tag{5}$$

where x_0 = 0 is set since $0 \le x \le 1$ and it is enough to examine only the right part of the Gaussian curve – the results for the left one could be obtained from the symmetry properties.

As (5) is too complex to be used in finding the centroids of the probability masses, Taylor series are used instead to the second order for simplicity:

$$p(x) = \sum_{i=0}^{\infty} \frac{p^{(i)}(0)}{i!} x^i \approx p(0) + p'(0) + \frac{p''(0)}{2} x^2, \tag{6}$$

$$p'(x) = -\frac{x}{\sqrt{2\pi}\sigma^3} e^{-\frac{x^2}{2\sigma^2}}, \tag{7}$$

$$p''(x) = \frac{1}{\sqrt{2\pi}\sigma}\left(\frac{x^2}{\sigma^4} - 1\right) e^{-\frac{x^2}{2\sigma^2}}, \tag{8}$$

and substituting (7) and (8) into (6) the approximate distribution is:

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma}\left(1 - \frac{1}{2}x^2\right). \tag{9}$$

The stationary values of the m_i are defined by the set of nonlinear equations obtained by transition to analog domain:

$$m_i = E\{x \mid x \in U_i\} = \frac{\int_{A_i}^{B_i} x p(x) dx}{\int_{A_i}^{B_i} p(x) dx} = \tag{10}$$
$$= \frac{24\sqrt{2\pi}\sigma(B_i^2 - A_i^2) - 3c(B_i^4 - A_i^4)}{48\sqrt{2\pi}\sigma(B_i - A_i) - 8(B_i^3 - A_i^3)}, \forall \mathrm{i}$$

The solution of such a set could be done in the following way [1]. Let us have:

$$z = [m_1, m_2, ..., m_k]^T. \tag{11}$$

The equation to be solved is of the form:

$$z = f(z). \tag{12}$$

The first approximation then is $z^{(0)}$ and every subsequent approximation for the root is found by:

$$z^{(s+1)} = f(z^{(s)}). \tag{13}$$

The increase of the computational complexity using (13) is nonlinear with the number of grid points growing. So in [1] an expedient way of calculating the root is given and it consists of defining the point density q_i around m_i as the inverse of the length of the Voronoi set – $q_i = [(m_{i+1} - m_{i-1})/2]^{-1}$. As a result of that q_i can be expressed in the form $const.[p(m_i)]^{\alpha}$. Then passing from m_i to m_j it is true:

$$\alpha = \frac{\log(m_{i+1} - m_{i-1}) - \log(m_{j+1} - m_{j-1})}{\log[p(m_j)] - \log[p(m_i)]}. \tag{14}$$

For improved accuracy more values of the m_i are needed as we shall see in the next section.

B. Finding the One-Dimensional SOM Distortion Measure with Finite Grids

The objective function of the SOM is given by [1]:

$$E = \sum_i \sum_j \int_{x \in V_i} h_{ij} \|x - m_j\|^2 p(x) dx, \tag{15}$$

where V_i is the Voronoi set around m_i and h_{ij} is defined as:

$$h_{ij} = \begin{cases} 1, & \text{if} \quad |\mathrm{i} - \mathrm{j}| < 2 \\ 0, & \text{otherwise} \end{cases} \tag{16}$$

and i and j run over all the values defining h_{ij}.

Then (15) becomes:

$$E = \sum_i \sum_j \int_{C_i}^{D_i} (x - m_j)^2 p(x) dx =$$
$$= \sum_i \sum_j \frac{1}{\sqrt{2\pi}\sigma}\Big[-\frac{(D_i^5 - C_i^5)}{10} + \tag{17}$$
$$+ \frac{m_j(D_i^4 - C_i^4)}{4} + \frac{(2 - m_j^2)(D_i^3 - C_i^3)}{6} -$$
$$- m_j(D_i^2 - C_i^2) + m_j^2(D_i - C_i) \Big] = \sum_i \sum_j E_{ij}.$$

where N_i is defined in (1) and the borders C_i and D_i of the Voronoi set V_i are:

$$\begin{aligned} C_1 &= 0, \\ C_i &= \frac{m_{i-1} + m_i}{2} \quad \text{for} \quad 2 \le \mathrm{i} \le \mathrm{k}, \\ \mathrm{D_i} &= \frac{m_i + m_{i+1}}{2} \quad \text{for} \quad 1 \le \mathrm{i} \le \mathrm{k}\text{-}1, \\ \mathrm{D_k} &= 1. \end{aligned} \tag{18}$$

From (17) and (18) the gradient of E could be found as:

$$\frac{\partial E}{\partial m_i} = \frac{\partial}{\partial m_i} \sum_{j \in N_{i-1}} E_{i-1,j} + \frac{\partial}{\partial m_i} \sum_{j \in N_i} E_{ij} + \\ + \frac{\partial}{\partial m_i} \sum_{j \in N_{i+1}} E_{i+1,j} \quad , \tag{19}$$

but here will not be given due to its long form and since the calculations are trivial.

To obtain minimal distortion for the SOM optimal values for m_i should be found. A way of doing this is using the gradient-descent method according to:

$$m_i(t+1) = m_i(t) - \lambda(t)(\partial E / \partial m_i \mid_t), \tag{20}$$

where the scalar factor $\lambda(t)$ is typically in the range from 0.001 to 0.01 but even larger values of the order of 10 are tolerable [1]. Here E is of the fifth order, so at least one minimum can be found. The tendency is with the growth of $\lambda(t)$ the found minimum to become global not depending of the initial values of m_i.

III. Computational Results

As a simulation environment we use Matlab® R2009B over MS® Windows® XP® Pro SP3.

First α from (14) is found for different number of grid points. The more m_i are used the more accurate are the results. For $i = 4$ and $j = k - 3$ assuring negligible border effects 10, 25, 50, and 100 grid points are used. The same experiment is done with three other probability functions – linear, quadratic and linear-quadratic in [1], so here a direct comparison can be made. The results are given in Table I.

TABLE I
THE DERIVED ALPHA FOR FOUR DIFFERENT DISTRIBUTIONS OF THE INPUT

Grid points	Exponent α			
	Linear, [1]	Quadratic, [1]	Linear-quadratic, [1]	Normal
10	0.5831	0.5845	0.5845	0.5850
25	0.5976	0.5982	0.5978	0.5980
50	0.5987	0.5991	0.5987	0.5980
100	0.5991	0.5994	0.5990	0.5992

It is clearly seen from Table 1 that even for such different distributions as linear and quadratic from one side and the normal on the other, the resulting α is almost constant and very close to 0.6. Graphically the results are given on Fig. 1.

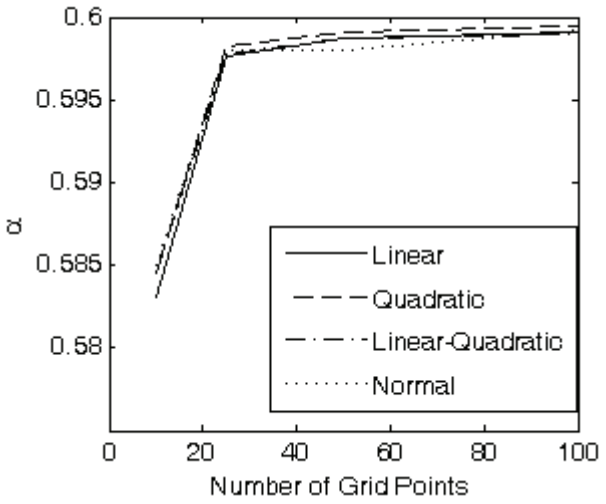

Fig. 1. Theoretically derived α as a function of the number of grid points for four different distributions of the input for the SOM

After computing the optimal m_i using (19) and (20) for the same number of grid points – from 10 to 100, it is possible to estimate α. Now when the minimal distortions have been assured this is done experimentally over real input data with quadratic and normal distribution around the winners. The results are given in Table II.

TABLE II
EXPERIMENTALLY ESTIMATED ALPHA FOR TWO DIFFERENT DISTRIBUTIONS OF THE INPUT AT THE MINIMAL DISTORTION MEASURE

Grid points	Exponent α	
	Quadratic, [1]	Normal
10	0.3281	0.2989
25	0.3331	0.3330
50	0.3333	0.3331
100	0.3331	0.3330

Obviously the values obtained here again do not depend on the number of grids and again are too close one to another despite the different form of the input distributions. The significant difference is with the results obtained by the theoretical derivations from (14). Now when we have the optimal m_i found it is seen that the exponent of the approximated state of the SOM is virtually equal to 1/3. This is actually a case coinciding with the optimal vector quantization [1] unlike the case of $\alpha = 0.6$ and confirms the correctness of the optimal m_i calculated. Graphically the results from Table II are given in Fig. 2.

In both cases, for $\alpha = 0.6$ and $\alpha = 0.3$, m_i can be considered as forming an elastic network expanding into the input feature space following the order of appearance of the consecutive

samples. The fact that the presented stochastic approximation (1) – (14) can not ensure the optimal case when $\alpha = 0.3$ does not mean that it is useless – actually saving the iterations from (20) and the preliminary fixation of m_i from (15)-(19) means considerable saving of computation time which is important in wide range of practical cases.

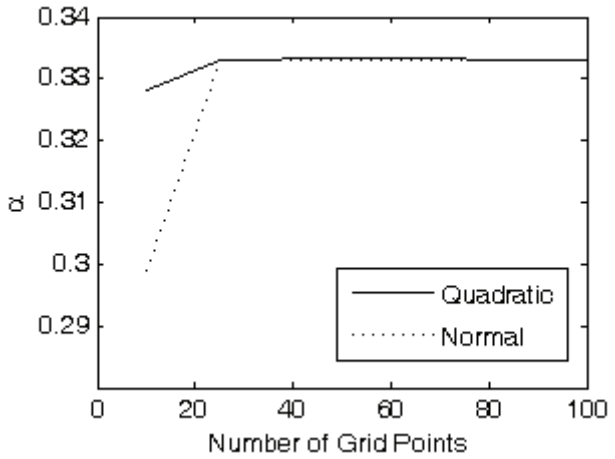

Fig. 2. Experimentally derived α as a function of the number of grid points for four different distributions of the input for the SOM at minimal distortion measure

One practical way of enhancing the work of a SOM is taking higher value for α, not only over 1 but even greater than 10, in the cases when the expanding network of nodes does not occupy at least half of the feature manifold. This step should be preceded by passing the training sample set in random order considerable number of times and only then should be used as an utter measure. In other cases proper action could be introducing normal distributed data into the area where the expansion is not wide enough and then to substitute it with the original one the number of epochs needed. In such situations the balance between the computational costs determined from (20) and the accuracy with its maximum defined by solving (19) should be carefully achieved.

IV. Conclusion

In this paper an approach for finding the stationary positions of the nodes of one-dimensional SOM has been presented in the case of normal density point input. The results are precise enough taking the advantage of very fast computation. Furthermore the distortion measure of the SOM using finite grid is calculated in the general case and it is shown that the positions of the nodes could be optimized iteratively at the cost of more computation time but which leads to results close to those from the optimal vector quantization.

The results achieved prove the correctness of the suggested approach which is considered useful in a large number of practical cases where the input data poses normal point density distribution.

Acknowledgement

This paper was supported by the National Fund for Scientific Research of the Bulgarian Ministry of Education and Science (Contract - BY-I-302/2007).

References

[1] T. Kohonen, *Self-Organizing Maps*, 3 ed., New York, Springer, 2001.

[2] H. Ritter and K. Schulten, “On the Stationary State of Kohonen's Self-Organizing Sensory Mapping”, Biological Cybernetics, Vol. 54, Issue 2, pp. 99-106, June 1986.

[3] H. Ritter, “Asymptotic Level Density for a Class of Vector Quantization Processes”, IEEE Transactions on Neural Networks, No. 2, pp. 173-175, 1991.

[4] D. Dersch and P. Tavan, “Asymptotic Level Density in Topological Feature Maps”, IEEE Transactions on Neural Networks, Vol. 6, No. 1, pp. 230-236, 1995.

[5] S. Luttrell, “Code Vector Density in Topographic Mappings: Scalar Case”, IEEE Transactions on Neural Networks, Vol. 2, No.4, pp. 427-436, 1991.

[6] R. Kountchev, D. Bojchev and D. Dobrev, “An Analysis of the Possibilities for Using Lattices for the Interpolation of the Exact Signal Vector Position of QAM Constellations”, In Proceedings of ICEST’2009, Veliko Tarnovo, Bulgaria, Vol.1, pp. 71-75, June 25-27, 2009.