SIGNAL PROCESSING VII: Theories and Applications
M. Holt, C. Cowan, P. Grant, W. Sandham (Eds.)

A Class of Robust Learning Vector Quantizers

R.YANG[1], M.GABBOUJ[2], I.PITAS[3], C.KOTROPOULOS[3]

[1] Nokia Research Center, Tampere, Finland, yang@research.nokia.fi

[2]Signal Processing Laboratory, TTKK, Finland, moncef@cs.tut.fi

[3] Dept.of Elec. Eng., Univ. of Thessaloniki, Greece, pitas@vergina.eng.auth.gr

Abstract. In this paper, we analyze a novel class of learning vector quantizers, namely weighted marginal median learning vector quantizer. The nonlinear based learning vector quantizer is robust against outliers and has better performance than the classical learning vector quantizer for some non-Gaussian distributions. It is shown that weighted median learning quantizer demonstrates similar convergence behavior to its linear counterpart. Simulations are carried to show that this type of nonlinear based learning vector quantizer will tend to approximate to the density function of the input vector. Its applications to color image quantization are considered and comparisons against its linear counterpart are made by simulations.

1 Introduction

Learning vector quantizer (LVQ) is an on-line learning algorithm which updates the codebook each time a training vector is presented and has been shown to yield equal or better results than the generalized Lloyd algorithm, the most widely used technique to design vector codebooks [1, 2, 3].

On the other hand, when training vectors do not obey multivariate Gaussian distribution, the linear operation in classical LVQ no longer yields the optimal estimation. Linear estimators have also poor performance against outliers. This motivated us to explore a class of weighted median based LVQ, where the input vectors are componentwise processed.

Median estimators have been shown to be optimal estimators for Laplacian distributions and are robust against outliers. Their properties in the filtering process have been studied extensively in past years [4, 5].

The Weighted Marginal Median LVQ (WMMLVQ) can be defined as follows [6].

In p-dimension space, the winner reference vector of a weighted marginal median LVQ is determined by

$$\mathbf{W}_c(n+1) = [x_{1wm}, x_{2wm}, \cdots, x_{pwm}]^T$$

where

$$x_{iwm} = \mathrm{MED}\{\, C_{i1} \diamond x_i(n), C_{i2} \diamond C_i(n-1), \ldots, C_{iN} \diamond x_i(0)\,\}. \quad (1)$$

Note that $N = n+1$ is the window size and $x_i(\cdot)$ is the set of the vector observations that have been assigned to each class. $(C_{i1}, C_{i2}, \cdots, C_{iN})$ is the coefficients for i-th reference vector.

2 Analysis

In the following, we shall use the *efficiency* ρ as the measure to compare the performance of classical LVQ and that of WMMLVQ.

The *efficiency* ρ is defined as the ratio of the variances of the classical LVQ and WMMLVQ. i.e.

$$\rho = \frac{\sigma^2{}_{CLVQ}}{\sigma^2{}_{WMMLVQ}}. \quad (2)$$

According to the statistical properties of WM operation, we have the following property.

Theorem: The *efficiency* ρ can be expressed as follows.

$$\rho = 4f^2(t_0)\sigma^2$$

where σ^2 is the input variance and $f(t)$, $F(t)$ is the input cdf and pdf, respectively,

$$t_0 = F^{-1}(\frac{1}{2}).$$

The theorem reveals some interesting facts about WMMLVQ. It shows that the performance of WMMLVQ is consistent with respect to that of the classical LVQ when both LVQs use the same weight vector. Specifically, we have the following observations about WMMLVQ.

Observation 1: Any sufficient and (or) necessary conditions which make the classical LVQ converge are valid for WMMLVQ.

Observation 2: The ρ is independent of the *schedules* (the adaptation steps) and is only the function of input distribution.

Here we list some common distributions ($\rho \geq 1$ indicates WMMLVQ is better):

For Gaussian distribution, $\rho = \frac{2}{\pi}$.

For Laplacian distribution, $\rho = 2$.

For the mixed distribution, i.e. the input pdf $f(t)$ can be expressed:

$$f(t) = \sum_{i=1}^{K} \epsilon_i f_i(t)$$

where

$$\sum_{i=1}^{K} \epsilon_i = 1$$

and $f_i(t)$ denotes the pdfs of various data classes. Denote m_i and $\sigma_i{}^2$ the mean and variance of the distribution $F_i(t)$, and m and σ^2 the mean and variance of input pdf $f(t)$. Then

$$m = \sum_{i=1}^{K} \epsilon_i m_i$$
$$\sigma^2 = \sum_{i=1}^{K} \epsilon_i \sigma_i{}^2 + \sum_{i=1}^{K} \epsilon_i m_i{}^2 - (\sum_{i=1}^{K} \epsilon m_i)^2$$

Suppose input classes are Gaussian distributions with same mean m_i, obviously

$$m = m_i, \qquad i = 1, \cdots, K$$

and $F(m) = \frac{1}{2}$,

$$f(m) = \frac{1}{\sqrt{2\pi}} \sum_{i=1}^{K} \frac{\epsilon_i}{\sigma_i}$$

The efficiency ρ

$$\rho = \frac{2}{\pi} \sum_{i=1}^{K} \epsilon_i \sigma_i{}^2 (\sum_{i=1}^{K} \frac{\epsilon_i}{\sigma_i})^2$$

When $K = 2$, WMMLVQ will be better ($\epsilon = 0.5$) if

$$\frac{\sigma_1{}^2}{\sigma_2{}^2} \geq 5$$
or
$$\frac{\sigma_1{}^2}{\sigma_2{}^2} \leq \frac{1}{5}$$

For two classes of gaussian distributions $N(m_1, \sigma_1{}^2)$ and $N(m_2, \sigma_2{}^2)$, the contaminated distribution is of the form

$$f(x) = \epsilon f_1(x) + (1 - \epsilon) f_2(x)$$

Suppose that $m_1 \leq m_2$. The decision region for class 1 is $x < d$ and for class 2 is $x > d$, where

$$d = (m_1 + m_2)/2.$$

The truncated distributions become

$$\begin{aligned} g_1(x) &= (ef_1(x) + (1 - e) f_2(x))/F_d, \\ g_2(x) &= (ef_1(x) + (1 - e) f_2(x))/(1 - F_d) \end{aligned} \qquad (3)$$

where

$$F_d = \int_{-\infty}^{d} f(x) dx.$$

In order to evaluate the overall performance of WMMLVQ, we introduce the *average efficiency* which is defined as

$$\bar{\rho} = \frac{\sum_{i=1}^{K} \sum_{j=1}^{p} \sigma^2{}_{ij}}{\sum_{i=1}^{K} \sum_{j=1}^{p} \sigma^2{}_{wmij}} \qquad (4)$$

In this case, the average efficiency $\bar{\rho}$ becomes

$$\bar{\rho} = \frac{4 g_1^2(m_{g1}) g_2{}^2(m_{g2})(\sigma^2{}_{g1} + \sigma^2{}_{g2})}{g^2{}_1(m_{g1}) + g^2{}_2(m_{g2})}.$$

The average efficiency is tabulated in Table 1 for the Gaussian distributions $f_1(x) = N(5, \sigma)$ and $f_2(x) = N(10, \sigma)$ with $\sigma = 1, 2, 3, 4, 5$ and $0 \leq \epsilon \leq 1$. Table 2 is for Laplacian distribution with means 5, 10 and variances from 1 to 5. From the tables we have the following observations:

Observation 1: WMMLVQ has better performance than the classical LVQ when the distribution is Laplacian, as we expected.

Observation 2: The average efficiency $\bar{\rho}$ is symmetric with respect to $\epsilon = 0.5$.

Observation 3: The average efficiency performs consistently with increasing variance for the Laplacian distributions, while for the Gaussian distributions the $\bar{\rho}$ is the best when $\sigma = 1$ and the worst when $\sigma = 2$, then it increases with the increasing of σ.

3 Simulations

As we know, the optimal VQ will approximate the density function in most practical applications. It has been demonstrated that the classical LVQ tend to approximate to the density function of the input

Table 1: The average efficiency for the Gaussian distribution

ϵ	$\sigma=1$	$\sigma=2$	$\sigma=3$	$\sigma=4$	$\sigma=5$
0	0.6269	0.5761	0.5743	0.5771	0.5797
0.05	0.6453	0.6498	0.6025	0.5882	0.5852
0.10	0.6419	0.5644	0.5970	0.5901	0.5865
0.15	0.6353	0.5102	0.5834	0.5851	0.5858
0.20	0.6313	0.4945	0.5663	0.5803	0.5834
0.25	0.6284	0.4938	0.5533	0.5734	0.5809
0.30	0.6266	0.4979	0.5431	0.5684	0.5782
0.35	0.6253	0.5227	0.5362	0.5639	0.5759
0.40	0.6245	0.5062	0.5325	0.5612	0.5740
0.45	0.6240	0.5087	0.5293	0.5593	0.5727
0.50	0.6239	0.5095	0.5285	0.5585	0.5727
0.55	0.6240	0.5089	0.5300	0.5586	0.5728
0.60	0.6244	0.5068	0.5319	0.5612	0.5741
0.65	0.6251	0.5031	0.5366	0.5640	0.5759
0.70	0.6263	0.4986	0.5436	0.5684	0.5776
0.75	0.6280	0.4943	0.5527	0.5743	0.5803
0.80	0.6307	0.4950	0.5670	0.5797	0.5836
0.85	0.6346	0.5102	0.5828	0.5862	0.5858
0.90	0.6409	0.5640	0.5985	0.5894	0.5868
0.95	0.6443	0.6492	0.6019	0.5893	0.5853
1.00	0.6298	0.5765	0.5737	0.5775	0.5798

Table 2: The average efficiency for the Laplacian distribution

ϵ	$\sigma=1$	$\sigma=2$	$\sigma=3$	$\sigma=4$	$\sigma=5$
0	1.3489	1.1549	1.0796	1.0470	1.0306
0.05	1.2219	0.7478	0.9503	1.0188	1.0393
0.10	1.5938	0.7936	0.9443	1.0390	1.0454
0.15	1.6876	0.9487	0.9996	1.0848	1.0497
0.20	1.7187	1.0842	1.0721	1.0904	1.0543
0.25	1.7483	1.1859	0.1415	1.0938	1.0549
0.30	1.7550	1.2606	1.1785	1.0954	1.0571
0.35	1.7632	1.3240	1.1800	1.0969	1.0570
0.40	1.7727	1.3684	1.1799	1.0974	1.0596
0.45	1.7960	1.3709	1.1794	1.0975	1.0592
0.50	1.7953	1.3684	1.1792	1.0994	1.0591
0.55	1.7957	1.3661	1.1795	1.0995	1.0593
0.60	1.7722	1.3686	1.1801	1.0976	1.0582
0.65	1.7624	1.3242	1.1803	1.0972	1.0587
0.70	1.7667	1.2652	1.1761	1.0958	1.0574
0.75	1.7467	1.1904	1.1422	1.0923	1.0554
0.80	1.7167	1.0851	1.0730	1.0891	1.0533
0.85	1.6850	0.9499	1.0028	1.0857	1.0503
0.90	1.5907	0.7952	0.9455	1.0414	1.0447
0.95	1.2194	0.7488	0.9507	1.0190	1.0387
1.00	1.3417	1.1548	1.0794	1.0467	1.0303

vectors in an orderly fashion [1]. In order to investigate the behaviors of WMMLVQ, we conducted the following simulation.

The input vectors were chosen to be two-dimensional and uniformly distributed in a square. At the beginning, each neural, which appears as a point in the square, was selected randomly in some area and all these neurals were connected by a lattice of lines which indicate correlations between these neurals. We used WMMLVQ to train these neurals by the input vectors. The results are shown in Fig. 1 and Fig. 2. Note that 900 neurals are used here.

From the simulations one may observe that WMMLVQs demonstrate similar behavior as their linear counterparts. They tend to approximate the density function of the input vectors, in this case the uniform distribution. It is expected that with more iterations the neural network will expand uniformly and finally reach the border of the square.

In order to demonstrate the performance of the weighted median based LVQ in image processing. We applied both the linear LVQ and WMMLVQ to color image quantization. The image used in the experiment is the RGB image "Lenna" with 512×480. 256 and 128 codewords were used. Both the linear LVQ and WMMLVQ were initialized by LBG algorithm. The figures of merit are shown in Table 3 and Table 4 for 256 codewords and 128 codewords, respectively. Note that training signals were collected from the image by the decimation factor in both directions.

By inspecting Table 3 and Table 4, it is seen that the weighted median LVQ achieves better results in most of cases, especially when the decimation factor decreases, compared to the linear LVQ. The best improvement of WMMLVQ over the linear LVQ is about 1.7 dB.

Table 3: Figures of merit, codewords=256

Decimation factor	Linear LVQ		WMMLVQ	
	PSNR (dB)	MAE	PSNR (dB)	MAE
1	31.41	9.37	32.61	7.83
2	31.85	8.91	32.71	7.78
4	32.37	8.37	32.70	7.80
8	32.66	7.95	32.52	7.94
16	32.16	8.27	31.95	8.46

Table 4: Figures of merit, codewords=128

Decimation factor	Linear LVQ		WMMLVQ	
	PSNR (dB)	MAE	PSNR (dB)	MAE
1	29.08	12.34	30.73	9.75
2	29.44	11.79	30.94	9.59
4	30.00	10.96	30.85	9.66
8	30.66	10.12	30.85	9.72
16	30.65	9.99	30.43	10.18

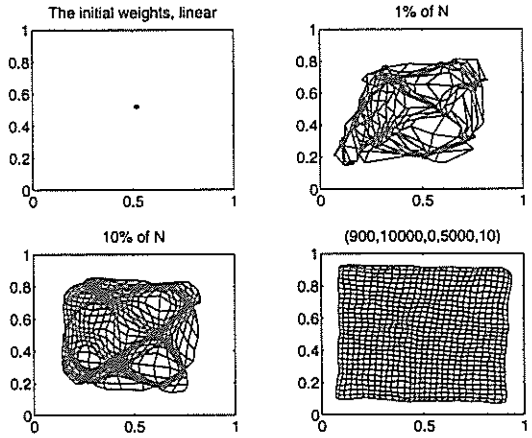

Figure 1: *The neural network during the order process for the classical LVQ*

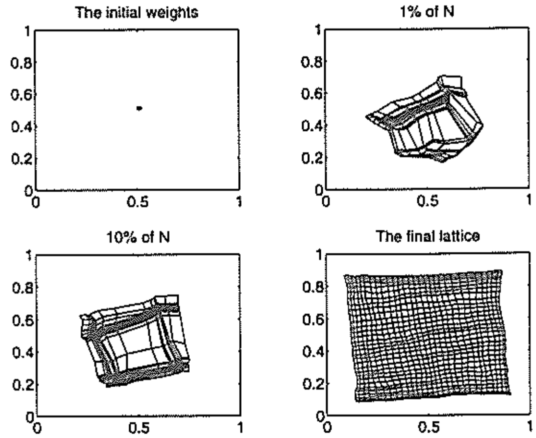

Figure 2: *The neural network during the order process for weighted marginal median LVQ, there are 900 neurals and the iteration number is 10000*

4 Conclusions

In this paper, we have introduced a novel class of learning vector quantizers which are based on weighted marginal median operations. The nonlinear based learning vector quantizer is robust against outliers and has better performance than the classical learning vector quantizer for some non-Gaussian distributions. It is shown by simulations that weighted median learning quantizer demonstrates similar convergence behavior to its linear counterpart, i.e., those nonlinear based learning vector quantizers will tend to approximate to the density function of the input vector. They were applied to color image quantization and comparisons against their linear counterparts are made by simulations.

References

[1] T. Kohonen, "The self-organizing Map", *Proc. of IEEE*, vol.78, no.9, Sept. 1990, pp.1464-1480.

[2] N.M. Nasrabadi and Y. Feng, "Vector quantization of images based upon the Kohonen self-organization feature maps," *Proc. 2nd ICNN Conf.*, vol.I, 1988, pp.101-105.

[3] P.-C. Chang, and R.M. Gray, "Gradient algorithms for designing predictive vector quantizers," *IEEE Trans. Acoust., Speech, Signal Processing*, vol.ASSP-34, pp.679-690, Aug. 1986.

[4] O. Yli-Harja, J. Astola, and Y. Neuvo, "Analysis of the properties of median and weighted median filters using threshold logic and stack filter representation," *IEEE on Acoust. Speech and Signal Processing*, vol.39, no.2, pp-395-410, Feb. 1991.

[5] I. Pitas, and A.N. Venetsanopoulos, *Nonlinear Digital Filters: Principles and Applications.* Hingham MA: Kluwer Academic Publishers, 1990.

[6] I. Pitas, C. Kotropoulos, N. Nikolaidis, R. Yang, and M. Gabbouj, "A class of order statistics learning vector quantizers," *Proc. of IEEE Int. Symposium on Circuits and Systems*, May 30-June 2, 1994, London, U.K.