ESANN'1997 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 16-17-18 April 1997, D-Facto public., ISBN 2-9600049-7-3, pp. 211-218

Kohonen maps versus vector quantization for data analysis

Eric de Bodt[1], Michel Verleysen[2], Marie Cottrell[3]

[1] Université catholique de Louvain, IAG-FIN, 1 pl. des Doyens,
B-1348 Louvain-la-Neuve, Belgium.
[2] Université catholique de Louvain, Electricity Dept., 3 pl. du Levant,
B-1348 Louvain-la-Neuve, Belgium.
[3] Université Paris I, SAMOS, 90 rue de Tolbiac,
F-75634 Paris Cedex 13, France

Abstract. Besides their topological properties, Kohonen maps are often used for vector quantization only. These auto-organised networks are often compared to other standard and/or adaptive vector quantization methods, and, according to the large literature on the subject, show either better or worst properties in terms of quantization, speed of convergence, approximation of probability densities, clustering,… The purpose of this paper is to define more precisely some commonly encountered problems, and to try to give some answers through well-known theoretical arguments or simulations on simple examples.

1. Introduction

Kohonen maps [1] are neural networks widely used in data analysis and vector quantization, and show two main characteristics : they realise a quantization of a continuous space, as other vector quantization techniques such as LBG [2], *k*-means [3],… and they exhibit a topological property, allowing to analyse the ordering of centroïds.

Recent literature seems to show that the VQ performances of Kohonen maps are worst than other techniques. As an example, here are some arguments recently developed:

- in [4], better clustering performances are shown with standard *k*-means algorithm;
- in [5], the same argument is developed to justify the use of VQ algorithm with MDS (Muldi-Dimensionnal Scaling) instead of Kohonen maps;
- in [6], the authors argue that there exists an exponent between the underlying density of vectors and centroids before and after VQ and propose to modify the algorithm to remedy to this « problem ». this non-unity exponent is known as the magnification coefficient [7]. Although different, this coefficient can be related to Lloyd's work [10].

We will try to see to which extend these arguments are pertinent regarding to definitions, well-known theoretical properties and simulations on simple examples.

2. Definitions and algorithms

A continuous space Ω, of dimension d, has a continuous probability density function (pdf) $f(x)$, and its cumulated density (or repartition function) is $F(x)=P(X<x)$, where P is the probability law, and the inequality is verified in each dimension. A vector quantization Φ is an application from a continuous space Ω, of dimension d, to a finite subset F (the "codebook") formed by n "codewords" $y_1, y_2,\ldots,y_n$ of Ω. The positions of the codewords are supposed to be fixed.

The aim of a vector quantization (VQ) is to compress the information by replacing all elements x of a cluster C_i (subset of Ω) by a unique centroid y_i. For a given number n of centroids, a good vector quantization minimises the loss of information, measured by the mean quadratic error:

$$\xi_n(f,\Phi) = \sum_{i=1}^{n} \int_{C_i} \left\| x - y_i \right\|^2 f(x)dx \qquad (1)$$

If N samples $x_1, x_2,\ldots,x_N$ are available (randomly chosen according to $f(x)$), this distortion is estimated by the intra-class variance:

$$\hat{\xi}(f,\Phi) = \frac{1}{N} \sum_{i=1}^{n} \sum_{x_j \in C_i} \left\| x_j - y_i \right\|^2 \qquad (2)$$

All classical VQ algorithms (LBG, k-means,…) minimise this distortion function by choosing appropriate centroids locations. There is no unique minimum of the distortion function, but we will not discuss this problem here.

The Kohonen algorithm [1] also realises a VQ. However, unlike more standard VQ algorithms, centroids are a priori ordered on a line (dimension 1) or a grid (dimension 2). The learning algorithm ensures that, after learning, "close" points in the input space Ω (according to an Euclidean distance measure for example) will be associated to the same centroid or to close centroids on the line or grid. This topological property is of most interest in many applications (see e.g. [9,10]), but this point will not be discussed here. We will focus our discussion on two aspects of the Kohonen algorithm : the vector quantization property, and the link between the distribution of centroids and the initial distribution $f(x)$.

During the learning phase, the Kohonen algorithm uses neighbourhoods of centroids taken into account during the adaptation; the size v of these neighbourhoods is usually decrased during learning. For v fixed, the algorithm is equivalent [11] to the minimisation of a generalised distortion function:

$$\xi_{n,v}(f,\Phi) = \sum_{i=1}^{n} \int_{\bigcup_{k \in V(i)} C_k} \left\| x - y_i \right\|^2 f(x)dx \qquad (3)$$

where $V(i)$ is the set of indexes in the neighbourhood of I, including i. This generalised distortion function can also be estimated through a finite set of samples x_1, $x_2,\ldots,x_N$, similarly to (2).

ESANN'1997 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 16-17-18 April 1997, D-Facto public., ISBN 2-9600049-7-3, pp. 211-218

3. Quantization

Concerning the quantization aspect of the Kohonen algorithm, it is clear that minimising $\xi_{n,v}$ is not equivalent to minimising ξ_n, and thus that it is hopeless to expect that the Kohonen algorithm with $v > 1$ neighbours could lead to a minimisation of distortion ξ_n. For example, we can mention the result in [12] there the difference between $\xi_{n,v}$ and ξ_n is measured when $\dim(\Omega) = 1$ and centroids are arranged on a line with 2 neighbours; it is shown that the order of magnitude of $\min(\xi_{n,v}) - \min(\xi_n)$ is C/n^3, where C is a constant.

In practical situations however, the learning phase of the Kohonen algorithm is usually concluded with $V(i) = \{i\}$ (the so-called "0-neighbour" case). In this case, $\xi_{n,v} = \xi_n$ and the Kohonen algorithm is strictly equivalent to classical VQ techniques. All previous steps (with neighbourhoods) can be seen as a "good" initialisation procedure of the VQ algorithm itself.

4. Density approximation

We study here the important following questions:
1) is it possible to estimate the initial distribution $f(x)$, using as only information the locations of centroids $y_1, y_2,\ldots,y_n$ (and the associated frequencies) ?
2) what is the relation between the discrete distribution of these n points and $f(x)$?

First, we have to mention a classical result (see [13] for example). If $f(x)$ is a continuous density function (with repartition function $F(x)$), and if N observations X_1, $X_2,\ldots,X_N$ are drawn independently according to $f(x)$, one can define an empirical repartition function $F_N(x)$ which converges uniformly to $F(x)$ when N goes to infinity:

$$F_N(x) = \frac{1}{N}\sum_{i=1}^{N} \mathrm{card}\left\{ X_i \,\middle|\, X_i < x \right\} \qquad (4)$$

This is equivalent to say that the empirical measure defined by points X_1, $X_2,\ldots,X_N$ converges (in law) to the initial probability P. The rate of convergence can also be characterised. Of course, when speaking about vector quantization, the n centroids y_1, $y_2,\ldots,y_n$ cannot be considered as independent drawings according to $f(x)$, and this result cannot be used.

In [14], the Kohonen algorithm terminated with 0 neighbour at the end of learning is studied. According to the comment at the end of section 3, the convergence of this algorithm is thus equivalent to the convergence of a classical VQ technique as the "competitive learning". The result in [14] shows that the centroids after VQ are a good discrete skeleton for reconstructing the initial density $f(x)$, provided that each centroid is weighted by the probability (estimated by the frequency) of its Voronoi region. In other terms, if $y_1, y_2,\ldots,y_n$ are the centroids after learning, and $C_1, C_2,\ldots,C_n$ the corresponding Voronoi regions, the following convergence (in law) is guaranteed:

$$\sum_{i=1}^{n} P(C_i)\delta_{y_i} \xrightarrow{law} P \qquad (5)$$

when n goes to infinity, and δ_{yi} is a Dirac function on y_i. This is equivalent to say that the empirical measure defined by centroids $y_1, y_2,\ldots,y_n$, weighted by the probabilities of the associated Voronoi regions, converges (in law) to the initial probability P.

Provided that centroids are adequately weighted, this results shows that it is possible to reconstruct the initial law, and the result is exact when the number of centroids goes to infinity. The authors of [14] also showed that the speed of convergence is better than with data obtained by independent random drawings.

Another interesting result can be found in [7], and has been completed by many results taking into account the number of neighbours used during learning, the shape of the neighbourhood function,… [15, 16]. We know that without weights, the initial distribution cannot be reconstructed exactly. More precisely, [7] shows that the best vector quantization (i.e. which leads to a minimisation of distortion ξ_n, without neighbour) corresponds to a discrete density $g(x)$ which converges asymptotically (when n goes to infinity) to:

$$g(x) = A\, f(x)^{d/d+2} \qquad (6)$$

where A is a constant. The Kohonen algorithm, when terminated with 0 neighbour, and other more classical VQ methods as the competitive learning, have this property. Nevertheless, it must be mentioned that it is not easy to visualise this property in simulations, because it concerns densities and is thus very sensitive to the observations. Another comment is that the exponent in (6) has no effect on *uniform* densities.

Although different, this result must be related to Lloyd's work on vector quantization [8], who also derives a similar relation between the original density $f(x)$ and the resulting $g(x)$. This last density is however defined in a different way in Lloyd's work (in a few words, $g(x)$ is here reconstructed as a staircase density in which the limits of intervals are defined by the centroids, and their weights are inversely proportional to their width. The probabilities of intervals are thus supposed equal, and the reconstructed density is shown to be related to $f(x)$ through an exponent as in (6). We insist on the fact that definitions of $g(x)$ are different in [7] and [8], and that further investigation would be necessary to compare these two results.

Finally, it must be mentioned that several authors [17, 6] proposed modifications of the algorithm to compensate the exponent in (6), and to obtain discrete densities of centroids closer from $f(x)$. To our opinion, the question is to know if the exponent in (6) must be considered as an undesired effect which must be compensated, or if the problem of reconstructing the initial density can be solved by using Pagès' result [14]. In practical situations, when the probabilities of Voronoi regions can be estimated by their frequencies (i.e. when the number of initial drawings is sufficiently large), Pagès'

result can be used, since it shows that the convergence to the initial density $f(x)$ is faster with centroids after VQ than with independent drawings on $f(x)$.

5. Simulations

We conducted a set of experiments on known simple one-dimensional pdfs, in order to verify if underlying pdfs before and after learning can be built to be asymptotically equal, without taking into account any exponent. The experiment is shown in Fig. 1 for uniform, gaussian, triangular and Poisson distributions respectively. Cumulated densities are build according to (4). We insist on the fact that this way of considering the pdf after VQ seems natural, but is different than for example in [6], where the locations of centroids *or* the winning frequencies are taken into account, but never together. Fig. 1 shows the cumulated probability densities before (plain line) and after (crosses) VQ, the differences between the two curves, and the Kolmogorov test. Both the differences and the Kolmogorov test clearly show that this way of approximating the underlying initial pdf is adequate.

We also conducted experiments on a 2-dimensional Gaussian distribution. Fig. 2 shows that the way to take into account locations of centroids and/or winning frequencies is determining to rebuild the initial underlying pdf. Fig. 2.a shows the theoretical distribution in 3-D, while Fig. 2.b presents the same distribution in 2-D (locations and sizes of the bubbles being strictly calculated according to the Gaussian distribution). Fig. 2.c and Fig. 2.d are build as presented in [6]. In the case of Fig. 2.c, only the positions of the units are taken into account. In the case of Fig. 2.d, bubbles are proportional to the winning frequencies of units but position are arbitrarily chosen. Fig. 2.e and f show, respectively in 2-D and 3-D, the reconstructed distribution using the Kohonen algorithm and (4). Results are self-speaking.

6. Conclusion

We shown theoretical results that could be used to reconstruct an unknown density $f(x)$ after vector quantization:

- if the probabilities of the Voronoi regions can be estimated, [14] ensures the convergence (in law) of the weighted empirical measure defined by the centroids to the initial distribution $f(x)$;
- [7] result shows the relation between the discrete density of the centroids (without weights) and $f(x)$; this relation can be inverted to reconstruct $f(x)$;
- several authors proposed modifications of the Kohonen algorithm to compensate the magnification coefficient, and thus to have a discrete density of centroids (without weights) converging to $f(x)$.

All these theoretical results concern asymptotic situations, when the number of centroids goes to infinity. Our simulations show that reconstructing the densities taking into account locations and frequencies (as suggested in [14]) is very efficient.

Whether one of the methods is more efficient or has a faster convergence to $f(x)$ when the number of centroids is finite remains an open question.

Finally, we recall that the Kohonen algorithm, terminated with 0-neighbour, can be seen as an initialization procedure of a classical VQ algorithm and thus, that the theoretical results after convergence are identical. Whether this initialization is better than other techniques also remains an open question.

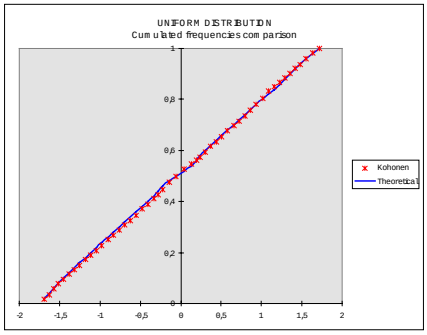
Acknowledgements

The authors are grateful to Anne Guérin-Dugué and Christian Jutten, from the Institut National Polytechnique de Grenoble, for fruitful discussions about the topics of this paper. Michel Verleysen is Research Associate of the Belgian National Fund for Scientific Research (FNRS).

References

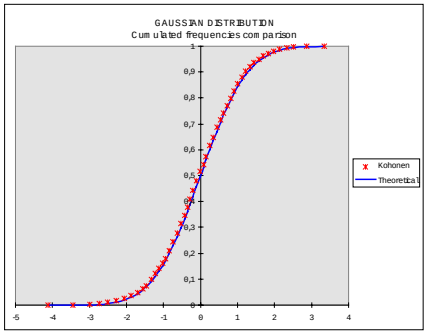
[1.] Kohonen T., *Self-organizing maps*, Springer, Berlin, 1995.
[2.] Linde Y., Buzo A., Gray R.M., *An algorithm for vector quantizer design*, IEEE Transactions on Communications, vol. COM-28, no. 1, January 1980.
[3.] MacQueen J. *Some methods for classification and analysis of multivariate observations*, Proc. Of the Fifth Berkeley Symposium on Math., Stat. and Prob., vol. 1, pp. 281-296, 1967.
[4.] Balakrishnan P.V. *A study of the classification capabilities of neural networks using unsupervised learning: a comparison with k-means clustering*, Psychometrika, vol. 59, no. 4, pp. 509-525, December 1994.
[5.] Flexer A., *Limitations of self-organizing maps for vector quantization and multidimensional scaling*, ftp ://ftp.ai.univie.ac.at/papers/oefai-tr-96-23.ps.z
[6.] Zheng Y., Greenleaf J.F., *The effect of concave and convex weight adjustments on self-organizing maps*, IEEE Transactions on Neural Networks, vol. 7, no. 1, pp. 87-96, January 1996.
[7.] Zador P.L., *Asymptotic Quantization Error of Continuous Signals and the Quantization Dimension*, IEEE Transactions on Information Theory, vol. IT-28, no. 2, pp. 139-149, March 1982.
[8.] Lloyd S.P., *Least squares quantization in PCM*, IEEE Transactions on Information Theory, vol. IT-28, no. 2, pp. 129-149, March 1982.
[9.] Cottrell M., Letremy P., *Classification et analyse des correspondances au moyen de l'algorithme de Kohonen : application à l'étude de données socio-économiques*, Proc. Neuro-Nîmes, 74-83, 1994.
[10.] Cottrell M., Ibbou S., *Multiple correspondence analysis of a crosstabulation matrix using the Kohonen algorithm*, Proc. ESANN'95, M.Verleysen Ed., Editions D Facto, Bruxelles, 27-32, 1995.
[11.] Ritter H., Martinetz and Shulten K., *Neural Computation and Self-Organizing Maps : an Introduction*, Addison-Wesley, Reading, 1992.
[12.] Fort J.C. and Pagès G., *Quantization versus Organization in the Kohonen S.O.M.*, Proc. ESANN'96, M.Verleysen Ed., Editions D Facto, Bruxelles, 85-89, 1996.
[13.] Bouleau N., *Variables Aléatoires et Simulation*, Hermann, 1986.
[14.] Pagès G., *Voronoï tesselation, Space quantization algorithms and numerical integration*, Proc. ESANN'93, M.Verleysen Ed., Editions D Facto, Bruxelles, 221-228, 1993.
[15.] Ritter H. and Shulten K., *On the Stationary State of Kohonen's Self-Organizing Sensory Mapping*, Biol. Cybern., 54, 99-106, 1986.
[16.] Dersch D. and Tavan P., *Asymptotic Level Density in Topological Feature Maps*, IEEE Transactions on Neural Networks, vol. NN-6, no. 1, pp. 230-236, January 1995.

[17.] Bauer H.-U., Der R. and Herrmann M., *Controlling the Magnification Factor of Self-Organizing Feature Maps*, Neural Computation, 8, 757-771, 1996.



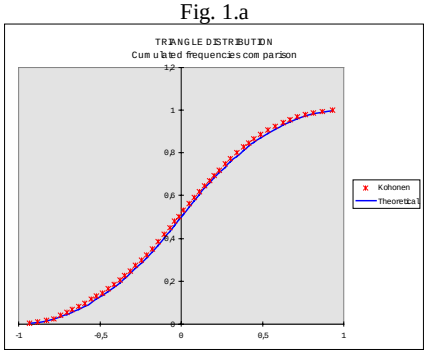
Kolomogorov Test : 0.01375 (CV : 0.193)

Fig. 1.a



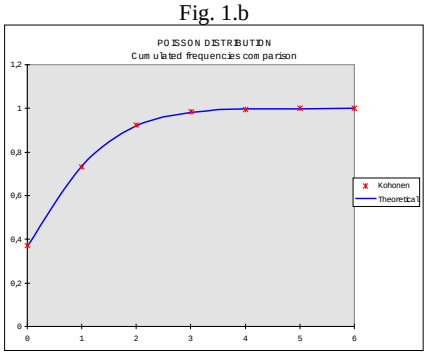
Kolomogorov Test : 0.02458 (CV : 0.193)

Fig. 1.b



Kolomogorov Test : 0.0194 (CV : 0.193)

Fig. 1.c



Kolomogorov Test : 0.0032 (CV : 0.387)

Fig. 1.d

Fig. 1 : 1-D Cumulated frequencies comparisons

ESANN'1997 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 16-17-18 April 1997, D-Facto public., ISBN 2-9600049-7-3, pp. 211-218

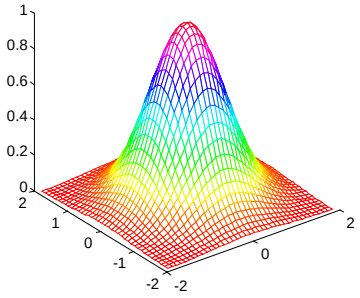


Fig. 2.a

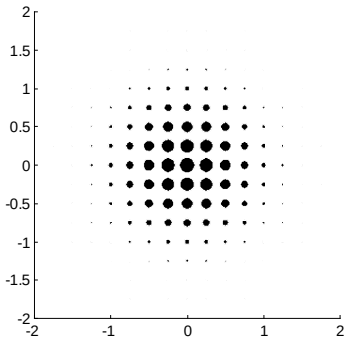


Fig. 2.b

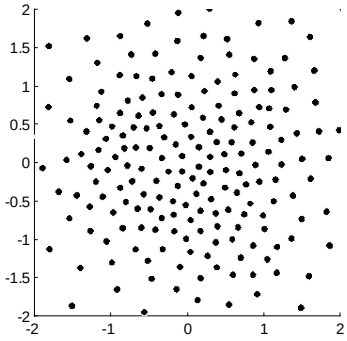


Fig. 2.c

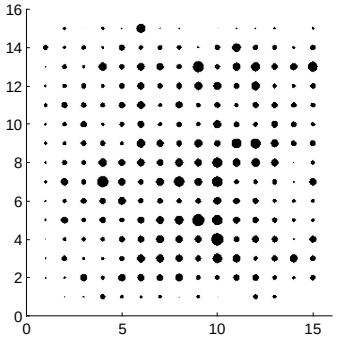


Fig. 2.d

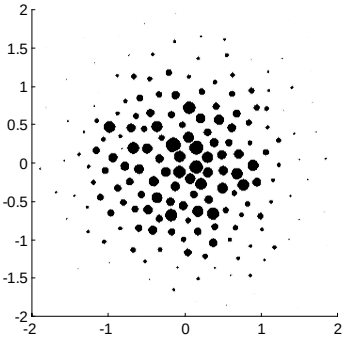


Fig. 2.e

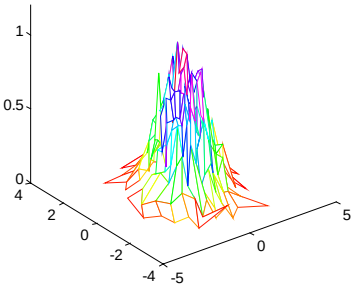


Fig. 2.f

Fig. 2 : 2-D simulations on Gaussian distribution (see text)