ESANN'1996 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 24-25-26 April 1996, D-Facto public., ISBN 2-9600049-6-5, pp. 103-110

A Kohonen Map Representation to Avoid Misleading Interpretations

Marie Cottrell* - Eric de Bodt+

*Centre de Recherche SAMOS - Université Paris 1 - Panthéon - Sorbonne
+Centre d'Etudes en Gestion Financière - Université Catholique de Louvain

Abstract. Recently, the Kohonen algorithm has been largely used to provide representations of large data sets and to complete the already extended range of data analysis techniques. In this paper, we address the problem of improving the graphical representation to avoid some classical difficulties of interpretation.

1. Using Kohonen Maps for Data Analysis

The Kohonen algorithm (Kohonen, 1982, 1989, 1995; Cottrell, Fort, 1987; Cottrell, Fort, Pagès, 1994) is a well-known unsupervised learning algorithm which produces an organized map composed of a fixed number of units.

Data analysis constitutes a classical field of applications for Kohonen maps. Oja (1982), Blayo and Demartines (1991), Cottrell, de Bodt and Henrion (1995) and others have shown that Kohonen maps can be perceived as a kind of principal component analysis, allowing the reduction of the number of dimensions of the studied data space and performing, in this sense, a projection (we refer to this application of Kohonen map as KACP). Cottrell and Letremy (1993) have shown that the Kohonen algorithm can be applied to contingency tables to implement a kind of Factorial Correspondence Analysis (Kouplet algorithm) and Cottrell and Ibbou (1995) have extended this idea to Multiple Correspondence Analysis (KMCA). Kohonen maps are also a quantification method. As mentioned above, at the end of the learning process, the weight vectors of the map units represent the mean profile of individuals which are (more or less) similar. The units of the map can therefore be considered as a set of prototypes representative of the total population. As such, Kohonen maps appear to be an unsupervised classification algorithm, where each unit represents a class (the interpretation is realized a posteriori by the observation of the weight vectors). Numerous applications have been made of these approaches.

However, drawing the Kohonen map after organization is not so obvious as it might seem at first sight. The classical square representation is only based on the physical ordering of the units. In the case of KACP, some authors put the number of individuals in each square (which gives the *frequency information*). In the case of Kouplet or KMCA, it is common practice to put the label of the individuals in the square. Based on the well-known feature of Kohonen maps of preserving the local topology of the input space, the map is then interpreted in terms of proximity of the

individuals (similar individuals tend to lie in adjacent units and vice-versa). This kind of interpretation may however be seriously misleading. The classical representation gives no idea of the local distances between the adjacent units and two neighbors can be, in the input space, very far from each other. In other words, the classical representation gives no idea of the eventual clusters that can exist on the map (Kohonen, 1995, p. 116). Two alternative approaches have been proposed to plot the Kohonen map.

- The first one is to use shades in a gray scale to put into light the clusters. The idea is the following : the closer two neighbor units are to each other, the darker is the shade chosen to plot them. The main problem with this representation lies in the fact that each unit has eight neighbors (except the border line units). Some authors (eg, Kohonen, 1995) suggest the use of the average of the distances between the unit and its neighbors. Others (eg, Kraaijveld, Mao and Jain, 1995) propose the use of the "maximum distance in the feature space of the corresponding unit to its four neighbors (East, West, North and South) in the network"[1]. While it is true that, if all the (chosen[2]) neighbors of a unit are close to that unit, a dark color will be used to plot it, this approach suffers from one main criticism. If, for example, in the set of the neighbors taken into account, one of the neighbors is far from the unit (in term of distance in the input space) while the other ones are close, using either criteria (mean or maximum), the chosen level of gray will be really misleading. Moreover, those kind of situations are impossible to detect without a closer look of all the local distances between all the neighbors on the map.

- An alternative proposal has been made by Demartines (1994). He introduces the "curvilinear" representation. The basic ideas are the following : for a 2-D Kohonen map, an arbitrarily chosen row of unit is designated as the horizontal axis and an arbitrarily chosen column will be the vertical one. The unit at the intersection of the two axes is plotted at point (0,0). The vertical (horizontal) coordinate of the units located on the horizontal (vertical) axis is 0. The vertical (horizontal) coordinate of a unit is then defined as the sum of the vertical (horizontal) coordinate of its bottom (left) neighbor and the distance, in the input space, between the considered unit and its bottom (left) neighbor. The representation obtained can also be seriously misleading. The diagonal distances between neighbor units on the representation are, in fact, only a geometric consequence of the cumulated horizontal and vertical distances and have no significance in terms of local proximity between the units.

---

[1] Kraaijveld, Mao and Jain, 1995, p. 551.

[2] We do not in fact understand why the authors limit the concept of the set of neighbors to the horizontal and vertical ones.

ESANN'1996 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 24-25-26 April 1996, D-Facto public., ISBN 2-9600049-6-5, pp. 103-110

2. A Kohonen Map representation that combines frequency and local distance information

The representation of the Kohonen map that we propose in this paper allows two kinds of information to be taken into account :

- the count of individuals for each unit of the map (the frequency information);
- the local distances between units.

The frequency information will be represented by a level of gray on the map. The darker the cell is, the higher is the number of individuals attached to the unit (a specific very light level of gray is chosen for a unit having no individuals attached to it).

The representation of the distances of a unit A to its neighbors is described in fig. 1 and derived as follows :

- The distances between all the pairs of neighbor units are calculated and the maximum Φ is determined as :

$$\Phi = \max_{i,j \text{ neighbors}} \left\| w_i - w_j \right\|$$

  where w_i and w_j are the weight vectors of units i and j of the Kohonen map.
  For the two-dimensional map (grid), we consider the 8 neighbors and for the one-dimensional map (string), we only consider 2 neighbors.
- For each neighbor B, an axis is built from the center of unit A in the direction of the center of unit B.
- For each neighbor B, on the axis, a point a is plotted proportionally to the distance between the two unit vectors w_A and w_B. More precisely :

$$Aa = AH\left(1 - \frac{\left\| w_A - w_B \right\|}{\Phi}\right)$$

  If units A and B are very close, a is approximately in H and conversely, if they are far from each other, a is near A.
- The same thing is repeated in the 8 directions and 8 points like a are found.
- The 8 points are connected to form an irregular octagon and the inside is colored as described above.

The result is a representation of the Kohonen map in which, in each square, each unit is represented by an octagon. The bigger it is, the closer the unit is to its neighbors (and vice-versa), so clusters appear to be regions in which octagons tend to be big and frontiers are regions largely unshaded. All the local distances can be compared because a common metrics has been used for all the units of the map (from 0 to the maximum of the local distances). In the next section, we will show three applications.

ESANN'1996 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 24-25-26 April 1996, D-Facto public., ISBN 2-9600049-6-5, pp. 103-110

It should be noted that the scale in the vertical and horizontal axes and in the diagonal axes are not the same. This introduces a slight visual distorsion. One way to overcome this difficulty could be to use a hexagonal lattice.

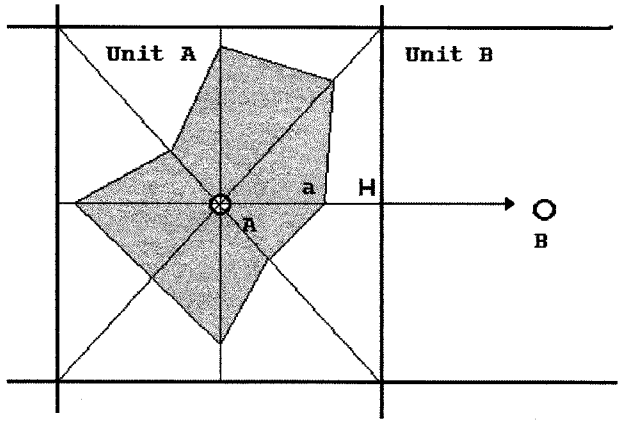

fig 1

3. Applications

To test the validity of the proposed representation, we have used three data sets[3]:

- the first one is composed of 15 simulated firms for which two financial ratios are computed : the Return On Investment (ROI) and the Total Debts on Total Assets (DTPT). Tab. 1 shows the data and fig. 2 demonstrates the three sub-groups (A,B,C) obtained. Fig. 3 presents the results of the projection on a one-dimensional Kohonen map using the approach described above and fig. 4, on a two-dimensional Kohonen map (in this case, without reduction of the number of dimensions). Fig. 3 clearly stresses that group B is farther from group C than group A, and that group C includes the greatest number of firms (unit 3 is the darkest). On fig. 4 and fig. 5, we see that group A is located in the left-bottom side of the Kohonen map, group C in the left-upper one and group B, in all the right one (the location of each unit is presented in fig. 5). The three areas tend to be separated by empty and distant units.
- the second one is the classical Iris data set (Fisher, 1936). Fig. 6 shows the results of the projection on a 4×4 Kohonen map. Results correspond to the a priori knowledge that we have about Iris data set. Setosa form a clearly distinct kind of Iris while Versicolor and Virginica are nearer to each other. These features are clearly shown up by our representation.

[3] Details on the implementation of the used Kohonen algorithm can be asked to the authors.

- the third one is composed of 9.659 Belgian firms for which 13 financial ratios have been computed on the basis of their 1991 financial statements (cf. tab. 4). The fig. 7 shows that the frequency information is well represented by our approach (cf. tab. 5), but it also stresses a difficulty met when using large real data sets. As underlined above, we use a common metrics for all the units which is based on the maximum of the local distances. If one of those distances is very large (in other words, if one individual is very atypical), the maximum is very large and therefore, all the other distances seem graphically to be very small. A solution to this problem could be, as proposed by one of the referees, the use of a logarithmic transformation of the distances.

We should like to express our thanks to Patrick Letremy and Ismaïl Ibbou for the "tips and tricks" that they provided to us for the implementation of the Kohonen algorithm and for their constructive advice.

References

- Blayo F., Demartines P. (1991), Data analysis : how to compare Kohonen neural networks to other techniques ?, In Proceedings of IWANN 91, Lectures Notes in Computer Science, Springer-Verlag, p. 469-476.
- Cottrell, M, Ibbou, S. (1995), *Multiple Correspondence Analysis of a crosstabulations matrix using the Kohonen algorithm*, in Proc of ESANN, M. Verleysen ED, D Facto, Bruxelles.
- Cottrell, M., de Bodt, E., Henrion, E-F (1995), Understandind the Leasing Decision with the Help of a Kohonen Map. An Empirical Study of the Belgian Market, submitted at the 1996 IEEE International Conference on Neural Networks.
- Cottrell, M., Fort, J.C., (1987), *Etude d'un algorithme d'auto-organisation*, Annales de l'Institut Poincaré, Vol. 23, 1, 1-20.
- Cottrell, M., Fort, J.C., Pagès, G. (1994), *Two or three things that we know about the Kohonen algorithm*, in Proc of ESANN, M. Verleysen ED., D Facto, Bruxelles.
- Cottrell, M., Letremy, P., Roy, E., (1993), *Analysing a Contingency Table with Kohonen Maps: a Factorial Correspondence Analysis*, Proceedings of IWANN'93, Springer Verlag, p. 305-311.
- Demartines, P. (1994), *Analyse de données par réseaux de neurones auto-organisés*, PHD Laboratoire TIRF, Institut National Polytechnique de Grenoble.
- Fisher, R.A. (1936), *The Use of Multiple Mesurements in Taxonomic Problems*, Annals of Eugenics, 7, 179-188.
- Kohonen T. (1982), Self organized formation of topologically correct feature maps, Biological Cybernetics, 43, p. 59.
- Kohonen, T. (1989), Self-organization and Associative Memory, 3°ed., Springer.
- Kohonen T. (1995), Self-Organizing Maps, Springer Series in Information Sciences Vol 30, Springer, Berlin.
- Kraaijveld M.A., Mao J. and Jain A.K. (1995), A Nonlinear Projection Method Based on Kohonen's Topology Preserving Maps, IEEE Transactions on Neural Networks, vol. 6, n°3, p. 548- 559.
- Oja E. (1982), A Simplified neuron model as a principal component analyzer, Journal of Mathematical Biology, vol. 15, p. 267-273.

- the third one is composed of 9.659 Belgian firms for which 13 financial ratios have been computed on the basis of their 1991 financial statements (cf. tab. 4). The fig. 7 shows that the frequency information is well represented by our approach (cf. tab. 5), but it also stresses a difficulty met when using large real data sets. As underlined above, we use a common metrics for all the units which is based on the maximum of the local distances. If one of those distances is very large (in other words, if one individual is very atypical), the maximum is very large and therefore, all the other distances seem graphically to be very small. A solution to this problem could be, as proposed by one of the referees, the use of a logarithmic transformation of the distances.

We should like to express our thanks to Patrick Letremy and Ismaïl Ibbou for the "tips and tricks" that they provided to us for the implementation of the Kohonen algorithm and for their constructive advice.

References

- Blayo F., Demartines P. (1991), Data analysis : how to compare Kohonen neural networks to other techniques ?, In Proceedings of IWANN 91, Lectures Notes in Computer Science, Springer-Verlag, p. 469-476.
- Cottrell, M, Ibbou, S. (1995), *Multiple Correspondence Analysis of a crosstabulations matrix using the Kohonen algorithm*, in Proc of ESANN, M. Verleysen ED, D Facto, Bruxelles.
- Cottrell, M., de Bodt, E., Henrion, E-F (1995), Understandind the Leasing Decision with the Help of a Kohonen Map. An Empirical Study of the Belgian Market, submitted at the 1996 IEEE International Conference on Neural Networks.
- Cottrell, M., Fort, J.C., (1987), *Etude d'un algorithme d'auto-organisation*, Annales de l'Institut Poincaré, Vol. 23, 1, 1-20.
- Cottrell, M., Fort, J.C., Pagès, G. (1994), *Two or three things that we know about the Kohonen algorithm*, in Proc of ESANN, M. Verleysen ED., D Facto, Bruxelles.
- Cottrell, M., Letremy, P., Roy, E., (1993), *Analysing a Contingency Table with Kohonen Maps: a Factorial Correspondence Analysis*, Proceedings of IWANN'93, Springer Verlag, p. 305-311.
- Demartines, P. (1994), *Analyse de données par réseaux de neurones auto-organisés*, PHD Laboratoire TIRF, Institut National Polytechnique de Grenoble.
- Fisher, R.A. (1936), *The Use of Multiple Mesurements in Taxonomic Problems*, Annals of Eugenics, 7, 179-188.
- Kohonen T. (1982), Self organized formation of topologically correct feature maps, Biological Cybernetics, 43, p. 59.
- Kohonen, T. (1989), Self-organization and Associative Memory, 3°ed., Springer.
- Kohonen T. (1995), Self-Organizing Maps, Springer Series in Information Sciences Vol 30, Springer, Berlin.
- Kraaijveld M.A., Mao J. and Jain A.K. (1995), A Nonlinear Projection Method Based on Kohonen's Topology Preserving Maps, IEEE Transactions on Neural Networks, vol. 6, n°3, p. 548- 559.
- Oja E. (1982), A Simplified neuron model as a principal component analyzer, Journal of Mathematical Biology, vol. 15, p. 267-273.

ESANN'1996 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 24-25-26 April 1996, D-Facto public., ISBN 2-9600049-6-5, pp. 103-110

ID	ROI	DTPT
1	-3	75
2	-2.4	77
3	-3.3	67
4	-2.7	70
5	17	45
6	18	48
7	20	46
8	18.5	52
9	19	51
10	4	11
11	5	15
12	4.5	4
13	5.7	7
14	3.9	9
15	5.1	8

tab. 1

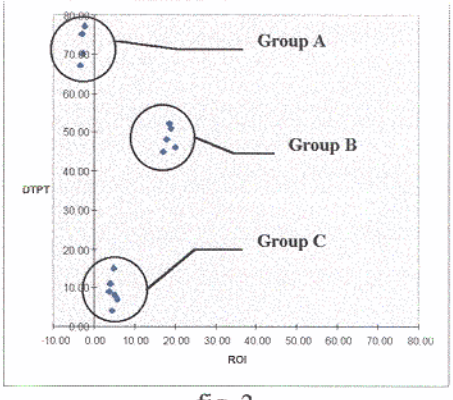

fig. 2

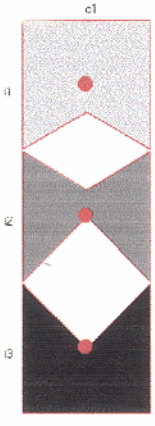

fig. 3

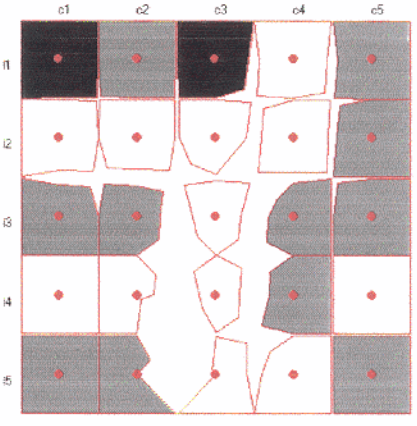

fig. 4

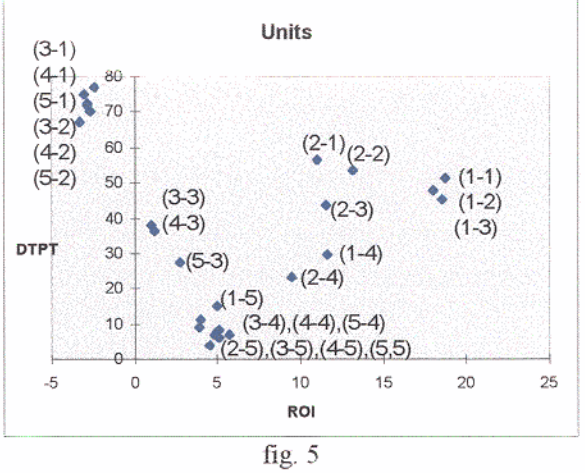

fig. 5

ESANN'1996 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 24-25-26 April 1996, D-Facto public., ISBN 2-9600049-6-5, pp. 103-110

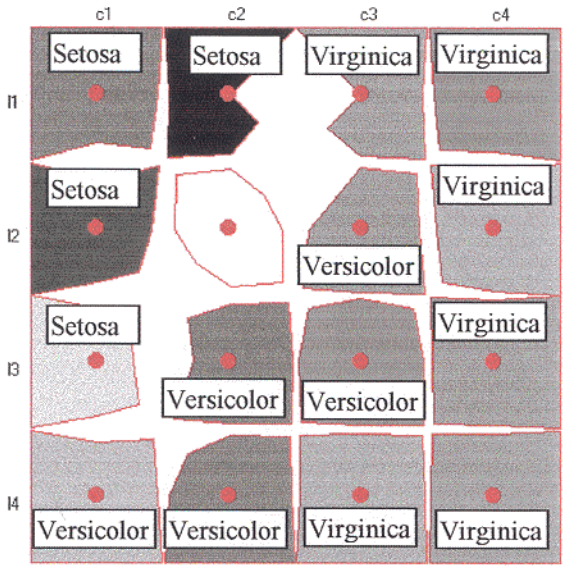

fig. 6

Count		Col				
Line	Indiv	1	2	3	4	Grand Total
1	Setosa	12	21	0	0	33
	Virginica	0	0	9	10	19
1 Total		12	21	9	10	52
2	Setosa	16	0	0	0	16
	Versicolor	0	0	9	0	9
	Virginica	0	0	0	7	7
2 Total		16	0	9	7	32
3	Setosa	1	0	0	0	1
	Versicolor	0	13	10	1	24
	Virginica	0	0	1	8	9
3 Total		1	13	11	9	34
4	Versicolor	4	11	2	0	17
	Virginica	1	1	5	8	15
4 Total		5	12	7	8	32
Grand Total		34	46	36	34	150

tab. 3

ESANN'1996 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 24-25-26 April 1996, D-Facto public., ISBN 2-9600049-6-5, pp. 103-110

Ratio	*Description*
ROE	Return on Equity
ROA	Return on Assets
REAE	Operating income on operating assets
TNPT	Cash position on total asset
CFFT	Cash flow on short term liabilities
BFRPT	Needs in working capital on total assets
FRNPT	Net working capital on total assets
DLTPT	Long term debt on total assets
DTPT	Total liabilities on total assets
VDAC	Cash on current assets
RRPTP	Retained earnings on total assets
EF	Financial debt on total assets
DFG	Financial mortgage on total assets

tab. 4

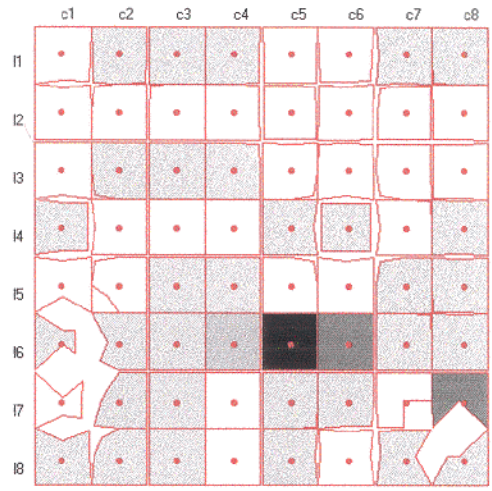

fig. 7

Count of Indiv	Col								
Line	1	2	3	4	5	6	7	8	Grand Total
1	3	9	216	269	6	4	48	182	737
2	151	9	152	7	5	3	3	7	337
3	47	590	666	65	17	216	148	3	1752
4	4	330	250	57	64	2	151	19	877
5	43	215	45	47	10	457	34	8	859
6	2	11	439	394	440	461	28	9	1784
7	1	798	405	128	91	30	21	783	2257
8	5	210	414	145	272	2	4	1	1053
Grand Total	256	2172	2587	1112	905	1175	437	1012	9656

tab. 5