ELSEVIER

May 1995

Pattern Recognition Letters 16 (1995) 507–515

Pattern Recognition Letters

Geometrical treatment and statistical modelling of the distribution of patterns in the n-dimensional Boolean space

Antônio de Pádua Braga *, Igor Aleksander

Imperial College, Department of Electrical and Electronics Engineering, Exhibition Road, London SW7 2BT, United Kingdom

Received 9 January 1994; revised 25 November 1994

Abstract

A method to estimate how the Boolean space is distributed in relation to two patterns separated by a distance h is presented. It allows the estimation of the overlap between the corresponding classes, which is a measure of correlation between them.

Keywords: Correlation of classes; Boolean space; Artificial neural networks

1. Introduction

In most cases in Pattern Recognition and Neural Networks fields there is no freedom to represent elements of the design and training sets as code words in order to minimize the overlapping among the classes determined by them. Elements of the design set are determined by the problem under study and can be located anywhere in E^n. Nevertheless, the knowledge of the amount of overlap between two arbitrary *r-spheres* (Thompson, 1983) is an important issue when working with Sparse Distributed Memory (Kanerva, 1984, 1988; Chou, 1989), Boolean Neural Networks (Braga, 1993, 1994; Aleksander, 1990), Pattern Recognition (Duda and Hart, 1973) and Coding Theory (Hamming, 1980). The overlap gives an idea of what is the percentage of the whole space that is *dubious* in relation to the centres of the *r-spheres*.

The aim of this paper is to present a model to estimate how the Boolean space is distributed in relation to two fixed patterns, which enables the prediction of the amount of overlap between the two corresponding *r-spheres*. The results enable a comparison with the previous work of Kanerva on the same subject (Kanerva, 1984, 1988). Such a comparison shows that for small values of r, the results of the two models are similar and quite accurate but, when r is large, the model presented here tends to be more precise when compared with computer results. Kanerva's model of Sparse Distributed Memory (SDM) uses small *r-spheres* and the expressions developed for that model are quite precise in the range of r demanded to prove the ideas behind SDM. The work presented here extends the one developed for SDMs in the sense that it is precise in the whole range of variation of r, being useful for SDMs as well as for any other model of Boolean Neural Networks and Pattern Recognition applications in general.

* Corresponding author.

SSDI 0167-8655(95)00120-0

2. The third side of the triangle

The problem addressed in this section is known as the "problem of the third side of the triangle" (Kanerva, 1984, 1988), due to the geometrical analogy which exists between a triangle in a plane and the relative distances among three arbitrary points in the Boolean space. It is important to note that Hamming instead of Euclidean distances are used as metric measures in this work. Although the rules of Euclidean Geometry cannot be directly applied to solve the problem, the geometrical analogy shown in Fig. 1 helps to visualise it.

It is aimed here to determine how the patterns that are at fixed distance r of ξ_1 are distributed in the space in relation to ξ_2. By the triangle and circle analogy of Fig. 1, it is suggestive that such a distribution starts from a minimum, increases to a maximum and then returns to the minimum in the same way that a binomial distribution does. It was observed that under certain circumstances the binomial approximation yields relatively good results, but there are other situations where this rule does not apply and the binomial approximation is poor. It is shown in the next sections that such a distribution resembles much more a hypergeometric than a binomial distribution and that the patterns at fixed distance of ξ_1 are seen by ξ_2 as clusters sparsely scattered in the space. Each cluster boundary defines a set of patterns that are at distance r from ξ_1 and distance δ from ξ_2.

3. The simplest case: $0 \leqslant r \leqslant h$ and $0 \leqslant r \leqslant (n-h)$

In this situation, the radius r in relation to ξ_1 is smaller than h and $(n-h)$. The procedure to obtain

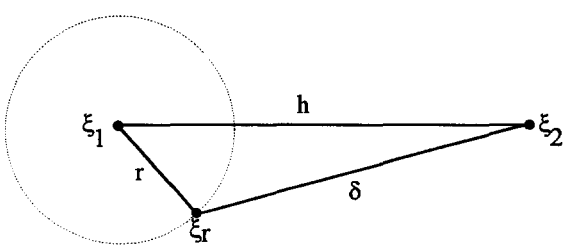

Fig. 1. Geometrical view of the "problem of the third side of the triangle". What is the distribution of δ?

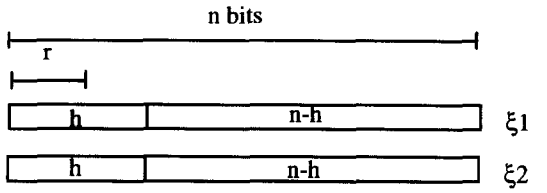

Fig. 2. Patterns ξ_1 and ξ_2 are separated by a distance h in the space $\{0, 1\}^n$.

a pattern at distance r from a fixed pattern ξ_1 is to invert r bits in its field of n bits. Using this procedure it is possible to obtain a maximum of $\binom{n}{r}$ different patterns at distance r from ξ_1. Fig. 2, where ξ_1 and ξ_2 are seen divided into two fields, facilitates the visualisation of this process in relation to ξ_2. All the bits in the *h-field* are different and all the bits in the *$(n-h)$-field* are equal when ξ_1 and ξ_2 are compared. Therefore, every pattern at distance r from ξ_1 should keep a known relationship with ξ_2 depending on whether the r bits were taken from the *h-field*, from the *$(n-h)$-field* or from both at the same time.

It can easily be seen that if all r bits are taken from the *h-field*, the patterns generated are at a distance $(h-r)$ from ξ_2 and if they are taken from the *$(n-h)$-field*, they shall be located at a distance $(h+r)$ from ξ_2. In the case they are taken from both fields at the same time, the distance δ is given by $h+r-2hr$, where hr is the number of bits taken from the *h-field*. In general, expression (1) gives the value of δ as a function of h, r and hr.

$$\delta = h + r - 2hr \tag{1}$$

Therefore, the number of patterns in each cluster at distance δ from ξ_2 is a function of the number of patterns which can be generated from ξ_1 after inverting hr bits in the *h-field* and $(r-hr)$ in the *$(n-h)$-field*. The size of each cluster can be calculated by the following expression:

$$T_r = \binom{n-h}{r-hr} \times \binom{h}{hr} \tag{2}$$

with hr varying from 0 to r for this particular case.

As it can be noticed from expression (2), the distribution of δ resembles a hypergeometric distribution. The main difference between the hypergeometric and the distribution of δ is that the second is

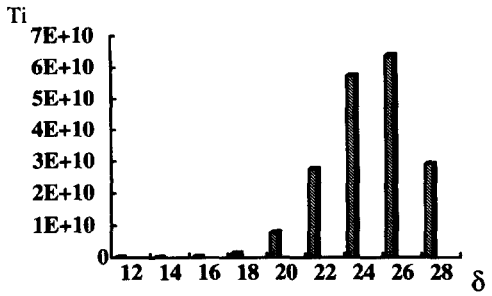

Fig. 3. Distribution of the clusters in relation to ξ_2 when $n = 100$, $h = 20$ and $r = 8$. It is bounded by the lower limit 12 and upper limit 28.

a discrete distribution with gaps between subsequent clusters, being bounded by the upper and lower limits of δ. For example, if $n = 100$, $h = 20$ and $r = 8$, it is obtained that the distribution of δ is formed by 9 clusters at distances 28, 26, 24, 22, 20, 18, 16, 14 and 12 and that each cluster has a different number of patterns as can be seen from Fig. 3. Such a distribution can also be visualized in terms of the triangle and circle analogy of Fig. 1, where each individual triangle has sides with sizes 20 (distance $h = 20$) and 8 (radius $r = 8$) with the "third side" being defined by one of the 9 clusters obtained above. Each cluster defines an individual triangle and the 9 triangles – one for each cluster – define the distribution of the clusters in the space. Therefore, the triangles mapped onto the plane have sides – in terms of Hamming distances: {20, 8, 28}, {20, 8, 26}, {20, 8, 24} and so on.

4. General solution

Based on the reasoning developed in the previous section, it is possible to find a general solution to obtain the distribution of δ when n, h and r are known. Expressions (1) and (2) are general solutions, but the bounds of hr and δ demand a more careful analysis in the different situations that may appear. Table 1 shows the values of the bounds of δ and hr for each possible relation among n, h and r. From a simple inspection in the table, it is possible to deduce that the lower bound of δ is defined by $|h - r|$ and its upper bound by $n - |n - (h + r)|$. From expression (1) the bounds of hr are then obtained:

$$hr_{\text{lower}} = \frac{h + r - n + |n - (h + r)|}{2}, \tag{3}$$

$$hr_{\text{upper}} = \frac{h + r - |h - r|}{2}. \tag{4}$$

The intersections of the functions $\phi_1(r) = |h - r|$ and $\phi_2(r) = n - |n - (h + r)|$ occur at $r = 0$ and $r = n$, which creates an envelope for the clusters within the range $0 \leqslant r \leqslant n$. A detailed description of the regions in the plane formed by the two functions can be seen in Fig. 4, together with an example for

Table 1
Bounds of δ and hr for every possible range of h and r

Range of h	Range of r	hr (lower)	hr (upper)	δ (lower)	δ (upper)
$h > n/2$	$0 \leqslant r \leqslant (n-h)$	0	r	$(h-r)$ $r < h$	$(h+r)$ $h+r < n$
$h > n/2$	$(n-h) < r \leqslant h$	$r-n+h$	r	$(h-r)$ $r < h$	$2n-(h+r)$ $h+r > n$
$h > n/2$	$h < r \leqslant n$	$r-n+h$	h	$-(h-r)$ $r > h$	$2n-(h+r)$ $h+r > n$
$h < n/2$	$0 < r \leqslant h$	0	r	$(h-r)$ $r < h$	$(h+r)$ $h+r < n$
$h < n/2$	$h < r \leqslant (n-h)$	0	h	$-(h-r)$ $r > h$	$(h+r)$ $h+r < n$
$h < n/2$	$(n-h) < r \leqslant n$	$r-n+h$	h	$-(h-r)$ $r > h$	$2n-(h+r)$ $h+r > n$
$h = n/2$	$0 < r \leqslant h$ $0 < r \leqslant (n-h)$	0	r	$(h-r)$ $r < h$	$(h+r)$ $h+r < n$
$h = n/2$	$h < r \leqslant n$ $(n-h) < r \leqslant n$	$r-n+h$	h	$-(h-r)$ $r > h$	$2n-(h+r)$ $h+r > n$

the case where $n = 8$ and $h = 3$. It is also worth to observe that any pattern at distance $r < h/2$ is considered to belong to class ξ_1 and any pattern at distance $r > (2n - h)/2$ is considered to belong to class ξ_2. Picking-up random patterns in the range $h/2 \leqslant r \leqslant (2n - h)/2$ can result in a classification in class ξ_1 or class ξ_2 with different expectations as a function of r. For example, considering Fig. 4, if a random pattern is picked-up at radius 2, there are 25 patterns which are closer to ξ_1 and 3 which are closer to ξ_2, among the total of 28 patterns. On the other hand, if a random pattern is picked-up at radius 5, there are only 16 patterns closer to ξ_1 among the total of 56. Therefore, the chance to pick-up a pattern closer to ξ_1 at radius 2 is $25/28$ and at radius 5 it decreases to $16/56$. Having a pattern recognition task in mind, it could be interpreted as if a random noise of 25% (radius 2 for this case) is added to ξ_1, the chance that the noisy pattern is classified as more similar to ξ_1 than to ξ_2 in terms of Hamming distance is $25/28$. The probability of the noisy pattern being classified as more similar to ξ_1 decreases with the increase of the amount of noise added to the original pattern, being zero for values of r greater than $(2n - h)/2$.

Fig. 5 shows all possible situations of $\delta(r)$ for the space $\{0, 1\}^8$, where a matrix was generated for each different value of h. The sizes and positions of the clusters were generated by counting the patterns using a computer program. The distributions generated by counting match those generated by expressions (1) to (4).

The observation of Fig. 5 also shows the change in the shape and size of the parallelogram formed by functions ϕ_1 and ϕ_2. The amount of the total area occupied by it gives an idea of the amount of clusters which are scattered in the space. The greater the area is, the greater is the number of clusters on the average and the greater is the spreading or variance of δ for fixed r. The normalized area can be calculated as follows:

$$a_n = \frac{\int_0^n n - |n - (h + r)| \, dr - \int_0^n |h - r| \, dr}{n^2}, \tag{5}$$

$$a_n = 2p(1 - p) \tag{6}$$

where $p = h/n$.

Function a_n has its maximum at the point $h = n/2$ ($p = 0.5$), where it occupies half the total area, as can be seen in Fig. 6. In other words, when the two patterns are orthogonal, the clusters are more sparsely distributed in the space. For large n it could be interpreted that when the patterns are orthogonal they are *indifferent* to each other, as pattern ξ_2 visualises the space at distance r of ξ_1 in the same way it does in relation to the entire space, or as a binomial distribution (Kanerva, 1984, 1988).

5. Amount of overlap

The clusters which are in the overlap of the classes defined by ξ_1 and ξ_2 are those located in the sites where $r = \delta$, or in the diagonal of the matrix. The observation of Fig. 5 suggests a binomial distribution for the terms in the diagonal as a function of r when h is even. It can be shown by making $\delta = r$ in expression (1) and taking it to (2):

$$T_r = \text{Constant} \times \binom{n - h}{r - h/2}. \tag{7}$$

As can be observed from Fig. 5, the terms in the overlap have a bell-shape, which suggests that it can be approximated using the standard form of the Normal Distribution Function (Feller, 1968) for large n. The best match found for such a distribution in relation to the total number of patterns in the overlap was the normal distribution with mean $n/2$ and standard deviation $\sqrt{n - h}/2$. Each term of such distribution is equivalent to $T_r/N_o(h)$, where $N_o(h)$

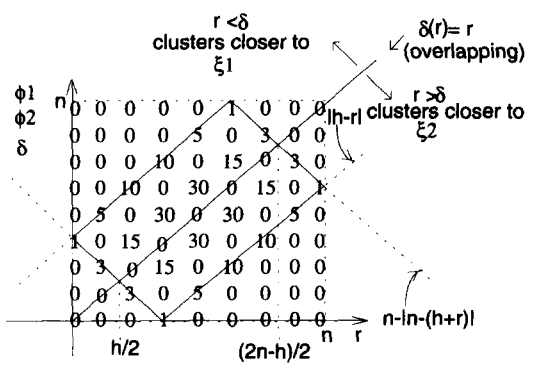

Fig. 4. Schematic view of the distribution of clusters as defined by expressions (1) to (4) for $n = 8$ and $h = 3$.

is the total number of patterns in the overlap for the distance h. Therefore:

$$\frac{T_r}{N_0(h)} = \frac{1}{\frac{1}{2}\sqrt{n-h}\sqrt{2\pi}} \exp - \frac{(r-n/2)^2}{2(\sqrt{n-h}/2)^2}. \quad (8)$$

The percentage of the total number of patterns that is in the overlap is

$$\varphi(h) = N_0(h)/2^n. \quad (9)$$

The function $N_0(h)$ can be obtained from expression (8) after substituting the hypergeometric distribution term T_r by its factorial form using Stirling's

h=0

δ \ r	0	1	2	3	4	5	6	7	8
8	0	0	0	0	0	0	0	0	**1**
7	0	0	0	0	0	0	0	**8**	0
6	0	0	0	0	0	0	**28**	0	0
5	0	0	0	0	0	**56**	0	0	0
4	0	0	0	0	**70**	0	0	0	0
3	0	0	0	**56**	0	0	0	0	0
2	0	0	**28**	0	0	0	0	0	0
1	0	**8**	0	0	0	0	0	0	0
0	**1**	0	0	0	0	0	0	0	0

h=1

δ \ r	0	1	2	3	4	5	6	7	8
8	0	0	0	0	0	0	0	**1**	0
7	0	0	0	0	0	0	**7**	0	**1**
6	0	0	0	0	0	**21**	0	**7**	0
5	0	0	0	0	**35**	0	**21**	0	0
4	0	0	0	**35**	0	**35**	0	0	0
3	0	0	**21**	0	**35**	0	0	0	0
2	0	**7**	0	**21**	0	0	0	0	0
1	**1**	0	**7**	0	0	0	0	0	0
0	0	**1**	0	0	0	0	0	0	0

h=2

δ \ r	0	1	2	3	4	5	6	7	8
8	0	0	0	0	0	0	**1**	0	0
7	0	0	0	0	0	**6**	0	**2**	0
6	0	0	0	0	**15**	0	**12**	0	**1**
5	0	0	0	**20**	0	**30**	0	**6**	0
4	0	0	**15**	0	**40**	0	**15**	0	0
3	0	**6**	0	**30**	0	**20**	0	0	0
2	1	0	**12**	0	**15**	0	0	0	0
1	0	**2**	0	**6**	0	0	0	0	0
0	0	0	**1**	0	0	0	0	0	0

h=3

δ \ r	0	1	2	3	4	5	6	7	8
8	0	0	0	0	0	**1**	0	0	0
7	0	0	0	0	**5**	0	**3**	0	0
6	0	0	0	**10**	0	**15**	0	**3**	0
5	0	0	**10**	0	**30**	0	**15**	0	**1**
4	0	**5**	0	**30**	0	**30**	0	**5**	0
3	**1**	0	**15**	0	**30**	0	**10**	0	0
2	0	**3**	0	**15**	0	**10**	0	0	0
1	0	0	**3**	0	**5**	0	0	0	0
0	0	0	0	**1**	0	0	0	0	0

h=4

δ \ r	0	1	2	3	4	5	6	7	8
8	0	0	0	0	**1**	0	0	0	0
7	0	0	0	**4**	0	**4**	0	0	0
6	0	0	**6**	0	**16**	0	**6**	0	0
5	0	**4**	0	**24**	0	**24**	0	**4**	0
4	**1**	0	**16**	0	**36**	0	**16**	0	**1**
3	0	**4**	0	**24**	0	**24**	0	**4**	0
2	0	0	**6**	0	**16**	0	**6**	0	0
1	0	0	0	**4**	0	**4**	0	0	0
0	0	0	0	0	**1**	0	0	0	0

h=5

δ \ r	0	1	2	3	4	5	6	7	8
8	0	0	0	**1**	0	0	0	0	0
7	0	0	**3**	0	**5**	0	0	0	0
6	0	**3**	0	**15**	0	**10**	0	0	0
5	**1**	0	**15**	0	**30**	0	**10**	0	0
4	0	**5**	0	**30**	0	**30**	0	**5**	0
3	0	0	**10**	0	**30**	0	**15**	0	**1**
2	0	0	0	**10**	0	**15**	0	**3**	0
1	0	0	0	0	**5**	0	**3**	0	0
0	0	0	0	0	0	**1**	0	0	0

h=6

δ \ r	0	1	2	3	4	5	6	7	8
8	0	0	**1**	0	0	0	0	0	0
7	0	**2**	0	**6**	0	0	0	0	0
6	**1**	0	**12**	0	**15**	0	0	0	0
5	0	**6**	0	**30**	0	**20**	0	0	0
4	0	0	**15**	0	**40**	0	**15**	0	0
3	0	0	0	**20**	0	**30**	0	**6**	0
2	0	0	0	0	**15**	0	**12**	0	**1**
1	0	0	0	0	0	**6**	0	**2**	0
0	0	0	0	0	0	0	**1**	0	0

h=7

δ \ r	0	1	2	3	4	5	6	7	8
8	0	**1**	0	0	0	0	0	0	0
7	**1**	0	**7**	0	0	0	0	0	0
6	0	**7**	0	**21**	0	0	0	0	0
5	0	0	**21**	0	**35**	0	0	0	0
4	0	0	0	**35**	0	**35**	0	0	0
3	0	0	0	0	**35**	0	**21**	0	0
2	0	0	0	0	0	**21**	0	**7**	0
1	0	0	0	0	0	0	**7**	0	**1**
0	0	0	0	0	0	0	0	**1**	0

h=8

δ \ r	0	1	2	3	4	5	6	7	8
8	**1**	0	0	0	0	0	0	0	0
7	0	**8**	0	0	0	0	0	0	0
6	0	0	**28**	0	0	0	0	0	0
5	0	0	0	**56**	0	0	0	0	0
4	0	0	0	0	**70**	0	0	0	0
3	0	0	0	0	0	**56**	0	0	0
2	0	0	0	0	0	0	**28**	0	0
1	0	0	0	0	0	0	0	**8**	0
0	0	0	0	0	0	0	0	0	**1**

Fig. 5. Distribution of the clusters at distance δ as a function of r and h. Elements $M(x,y)$ hold the number of patterns at distance y from ξ_2 and radius x from ξ_1.

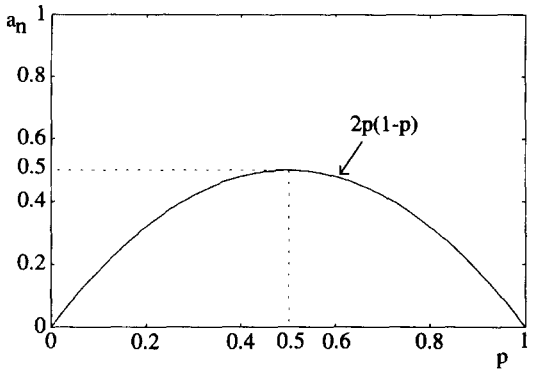

Fig. 6. Normalized area a_n is maximum when $h = n/2$.

formula (Feller, 1968) and making $r = n/2$ in both sides of the expression. The expression for $N_0(h)$ obtained in this way is then taken to expression (9), from which the function $\varphi(h)$ shown in expression (10) is finally obtained:

$$\varphi(h) = \begin{cases} 1 & h = 0, \\ \sqrt{2/\pi h} & h \geqslant 1. \end{cases} \tag{10}$$

Therefore, expression (10) determines the percentage of the whole space that is at the same distance from two arbitrary patterns ξ_1 and ξ_2 which are separated by the even distance h (if h is odd $\rightarrow$ $\varphi(h) = 0$). It is quite useful, as it provides a closed form to assess the amount of *dubious* patterns in the space as a function of the distance h. It is also significant to observe that the percentage of overlap does not depend on the number of dimensions n of the space, but only on the distance between the patterns. For example, if two patterns are at distance 14 in the space $\{0, 1\}^{20}$, there is a percentage of approximately 21% of the total number of patterns that has the same distance from them. For the particular case of the space $\{0, 1\}^{20}$, it corresponds to 223602 patterns of the total amount of 2^{20} possible. Fig. 7 shows the graphics of $\varphi(h)$ compared with experimental results for $n = 20$. For spheres with arbitrary radius r_1, the percentage of patterns $\theta(r_1)$ in the overlap can be obtained from the previous expressions:

$$\theta(r_1) = \frac{2}{\pi\sqrt{h(n-h)}} \int_{-\infty}^{r_1} \exp - 2\frac{(r - n/2)^2}{n-h}\, dr. \tag{11}$$

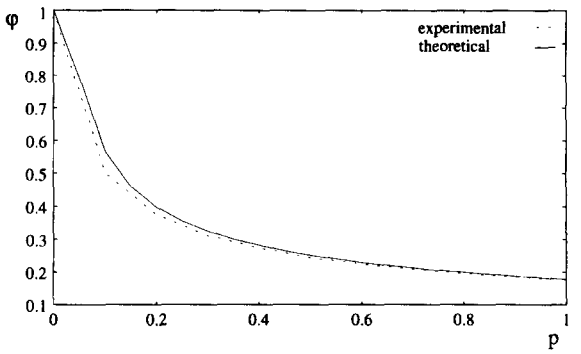

Fig. 7. Comparison between experimental and theoretical results of $\varphi(h)$ for $n = 20$ ($p = h/n$).

Kanerva (1984, 1988) developed a similar expression to estimate the amount of overlap, which is:

$$g(p, u) = \int_{h/n}^{1} \frac{1}{2\pi\sqrt{u(1-u)}} \exp - \frac{1}{2}\frac{C_p^2}{1-u}\, du \tag{12}$$

where $C_p = (r_p - n/2)/\sqrt{n/4}$.

As pointed out by Kanerva (1984, 1988), expression (12) is precise for small values of r, as can be observed in Fig. 8, where a comparison with expression (11) is made. For each point of the graphics in Fig. 8 there was calculated the average of the absolute errors of the two expressions in relation to values generated by counting in a computer program, having h varying from 0 to n for each value of r. It can be seen that expression (12) is very precise for all the values of r and can be used as a general

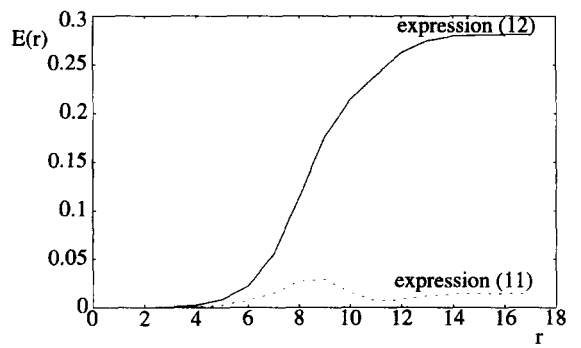

Fig. 8. Mean error of expressions (11) and (12) in relation to experimental results for $n = 20$.

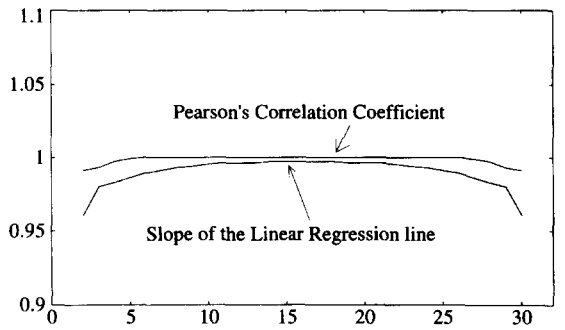

Fig. 9. Comparison between discrete distribution given by expressions (1) to (4) after cubic splines interpolation and the normal approximation of the hypergeometric distribution given by expressions (13) and (14).

solution for whatever value r assumes. The difference between the results of the two expressions persists for larger values of n.

6. Continuous approximations of the discrete distribution

In some situations, it may not be practical to determine the distribution of the clusters using an algorithmic procedure as the one presented in Sections 2 and 3. Despite the fact that the distribution of δ is actually discrete, an approximation by a continuous function can yield satisfactory results in some situations. Expressions (1) to (4) suggest that the distribution of δ for fixed r could be approximated by a hypergeometric distribution. The corresponding distribution that is defined by expression (2) has mean defined by expression (13), which was obtained by shifting the standard form of the mean of the hypergeometric distribution (Mendenhall and Sincich, 1988) by the function $\phi_1(r) = |h - r|$. The standard deviation adopted is shown in expression (14), which is in the standard form of the equivalent

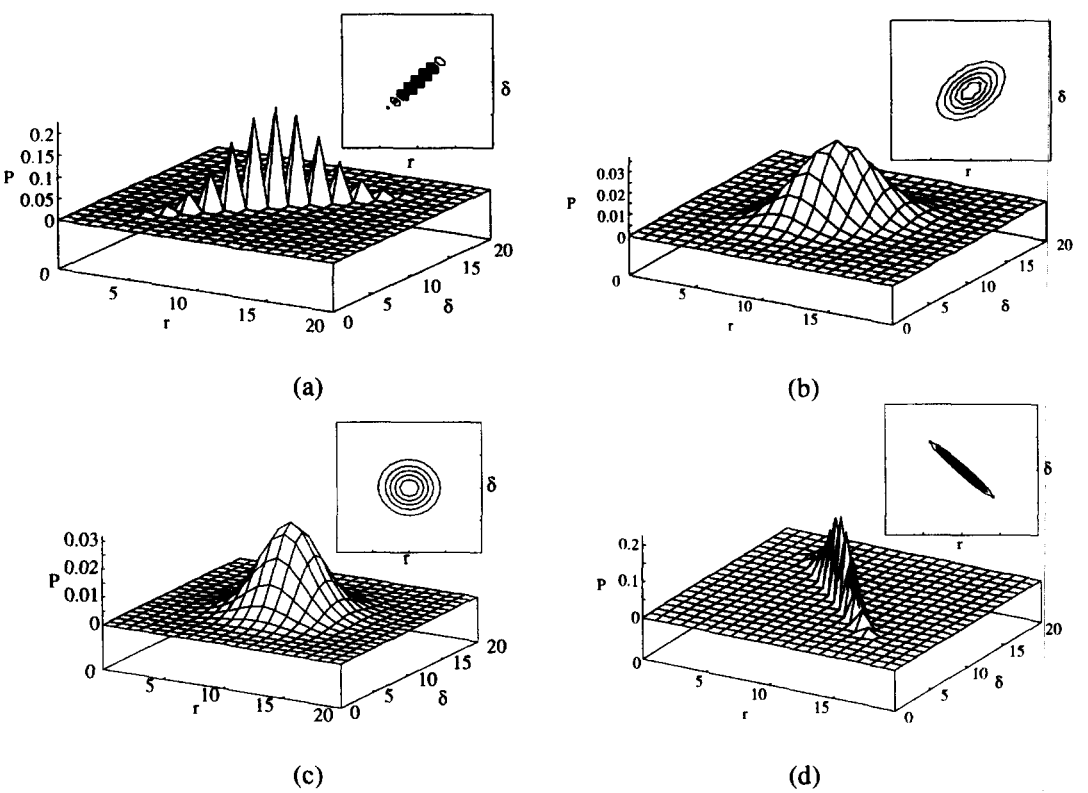

Fig. 10. Changes in the distributions of patterns as a function of the correlation coefficient for $n = 20$. (a) $\rho = 0.98$ ($h = 0$); (b) $\rho = 0.49$ ($h = 5$); (c) $\rho = 0$ ($h = 10$); (d) $\rho = -0.98$ ($h = 20$).

hypergeometric distribution (Mendenhall and Sincich, 1988).

$$m_h = \frac{h}{2} - r\left(\frac{h}{n} - \frac{1}{2}\right) \tag{13}$$

$$\sigma_h = \sqrt{\frac{h(n-h)r(n-r)}{n^2(n-1)}} \tag{14}$$

Such a hypergeometric distribution can be approximated by a normal one with the same mean and standard deviation for large n. Fig. 9 shows, for $n = 32$ and $h = 16$, a comparison between the normal approximation of the hypergeometric distribution given by expressions (13) and (14) and the corresponding discrete distributions generated by expressions (1) to (4) after interpolation using cubic splines. The normal approximation is excellent, as can be seen by the slope of the linear regression line and the Pearson's correlation coefficient shown in the graphics of Fig. 9. Both parameters are very close to 1, which means that the relationship between the two functions is strongly linear and that the numerical results of them are very close.

The change in the matrices of Fig. 5 also suggests that the distribution of the space (all the clusters present at different r's and δ's) in relation to the patterns ξ_1 and ξ_2 can be modelled by a bivariate normal distribution (Mendenhall and Sincich, 1988) with its correlation coefficient changing as a function of the distance between them. The best match for such a correlation coefficient is shown in expression (15).

$$\rho(h) = 0.98 - 1.96\frac{h}{n} \quad \begin{cases} 0 \leqslant h \leqslant n \\ -1 \leqslant \rho(h) \leqslant +1 \end{cases} \tag{15}$$

The means and standard deviations of the corresponding marginal distributions which fitted better with such approximation were, respectively, $m_1 = m_2 = n/2$ and $\sigma_1 = \sigma_2 = \sqrt{n/4}$. Some graphics of the surfaces generated by the bivariate normal distribution are shown in Fig. 10 to provide a comparison with Fig. 5. It can be observed that the form of the graphics of Fig. 10 changes as a function of the correlation coefficient between the two random variables r and δ (which is a function of h) in the same way that the matrices of Fig. 5 change as a function of the distance between the two patterns. The bivariate normal distribution has been successfully used to model the distribution of patterns in the Boolean space.

7. Conclusions

This paper presented a general solution for determining how the n-dimensional space is distributed in relation to two arbitrary patterns which are separated by a fixed distance in such a space. It was shown that such a distribution is discrete with gaps between its subsequent elements. The discrete distribution can be determined by using the expressions and procedures presented in Sections 2 and 3, which involves calculating its bounds, determining where the actual gaps are located and calculating each element of the distribution. In this way, it is possible to estimate how the space is divided into the classes characterized by the two patterns.

The procedures developed to estimate the discrete distribution allowed the deduction of expression (10), which is a closed form to determine the percentage of patterns in the overlap between the two classes. The closer are the patterns, the greater is the correlation between them and the greater is the overlap between the two classes. Therefore, the amount of overlap between the n-spheres (spheres of radius n) given by expression (11) can be seen as a measure of the correlation between the *classes* determined by the two patterns.

The geometrical treatment and statistical modelling by continuous functions presented in this paper provide the basic framework to determine the distribution of the Boolean space in relation to two fixed patterns. Such knowledge of the distribution of the space is an important issue when working with Artificial Neural Networks, Associative Memories, Pattern Recognition and Coding Theory.

Acknowledgements

Antônio de Pádua Braga would like to thank Universidade Federal de Minas Gerais and CNPq (grant number 202286/91-6), Brazil, for the joint support to this work.

References

Aleksander, I. (1990). Neural Systems Engineering: towards a unified design discipline? *IEE Computing and Control Engineering J.* 6 (1), 259–265.

Braga, A.P. (1993). On the information capacity of auto-associative RAM-based neural networks *Proc. ICANN-93*, Amsterdam, 13–16 Sept. 1993.

Braga, A.P. (1994). Predicting contradictions in the storage process of recurrent Boolean neural networks. *IEE Electronics Lett.* 30 (1), 55–56.

Chou, P.A. (1989). The capacity of the Kanerva associative memory. *IEEE Trans. Inform. Theory* 35 (2), 281–298.

Duda, R. and P. Hart (1973). *Pattern Classification and Scene Analysis.* Wiley, New York.

Feller, W. (1968). *An Introduction to Probability Theory and Its Applications*, 3rd edition. Wiley, New York.

Hamming, R.W. (1980). *Coding and Information Theory*, 2nd edition. Prentice-Hall, Englewood Cliffs, NJ.

Kanerva, P. (1984). Self-propagating Search: A Unified Theory of Memory. PhD Dissertation: Stanford University.

Kanerva, P. (1988). *Sparse Distributed Memory.* MIT Press, Cambridge, MA.

Mendenhall, W. and T. Sincich (1988). *Statistics for the Engineering and Computer Sciences*, 2nd edition. Dellen Publ. Co., San Francisco.

Thompson, T.M. (1983). *From Error-correcting Codes through Sphere Packing to Simple Groups*, The Carus Mathematical Monographs. The Mathematical Association of America.