Proceedings of the International Multiconference on Computer Science and Information Technology pp. 103–109

ISSN 1896-7094

Linear Ordering of Observation Space for Pattern Recognition

Juliusz L. Kulikowski[1]

Institute of Biocybernetics and Biomedical Engineering PAS,
4 Ks. Trojdena Str. 02-109 Warsaw, Poland

Abstract. The problem of linear ordering of observation space as a novel approach to pattern recognition based on non-parametric, combinatorial statistical tests is presented. The problem consists in ordering the elements of a discrete multi-dimensional observation space along a curve such that the elements belonging to different similarity classes are close each to each other as much as possible, the similarity classes are mutually separated and the length of the curve is minimized. Such a problem is, in general, NP-difficult and it is shown how its approximate solution can be reached. Its effective solution leads to a construction of pattern recognition algorithm based on combinatorial (serial) statistical tests.

Keywords: image interpretation, ontological models, hyper-relations, medical imaging.

1 Introduction

Most of the commonly known pattern recognition methods based on metrical approach (we differentiate such approach and the one based on artificial neural networks) use geometrical pattern recognition models having the form of hyper-planes separating the subsets of points in a multi-dimensional space, representing similarity classes of objects under observation [1,2,3]. Such models are particularly useful for examination of non-supervised pattern recognition methods based on Bayesian, correlation-based, etc., as well as of supervised pattern recognition learning methods (potential functions, *k-nearest neighbors (k-NN)* based, etc.), assuming that we have to do with well mathematically defined, homogenous, metric observation spaces. This condition is satisfied if the objects are represented by vectors of components expressed by the same physical units.

Otherwise, if the elements of observation space represent various object parameters it arises a problem of scaling. For example, it is not clear what is the sense of an Euclidean distance between two vectors used in medical diagnosis whose components are: x_1—patient's age [years], x_2 —patient's weight [kg], x_3—systolic blood pressure [hPa], x_4—diastolic blood pressure [hPa], etc. It also arises a question, is it better to express the blood pressure in hPa or in mm Hg, because their relative influence on the distance measure between vectors depends significantly on the units. In addition, it may happen that some Boolean components should be taken into consideration, like: x_5—was the patient hospitalized due to cardiac infarct [YES, NO]. It is evident that in such case (which in computer-aided medical diagnosis is rather typical) an observation space $\boldsymbol{X}$ consisting of vectors like $\mathbf{x} = [x_1, x_2, x_3, x_4, x_5]$ does not satisfy the metric space conditions. Even if normalization of components is used, it remains the problem of arbitrary choosing the normalization coefficients and of their influence on the results of pattern recognition. Pattern recognition methods neglecting this problem lead to algorithms which can be only locally used, their decisions made under different assumptions being, in fact, incomparable.

Pattern recognition methods based on the concept of linear ordered observation space make possible overcoming the above-mentioned difficulties. They originated with an attempt to use some non-parametric statistical tests to pattern recognition under strongly limited primary information about the similarity classes. It has been found that combinatorial (serial) tests satisfy well the conditions [4,5].

The idea can be easily geometrically illustrated. In Fig. 1 a 2-dimensional observation space is shown and three similarity classes of objects represented by training sets: S_1 whose elements are denoted by **x**, S_2 denoted by • , and S_3 denoted by ▬ . Unclassified elements have been denoted by **o** while **?** denotes a new-observed element being to be recognized.

If a new element **?** is to be recognized then, using the *k-NN* approach and a Manhattan metric, its 8-connective neighborhood should be examined. It contains: 3 unclassified elements **o**, 1 element **x**, 3 elements ▬ and 1 element •. Therefore, (for $k < 4$) **?** will be recognized as an element of S_3 (i.e. as ▬).

For serial statistical test using a linear ordering of observation space should be introduced. This can be reached in various ways, one of them being based on lexicographical ordering of elements: 1^{st} in rows,

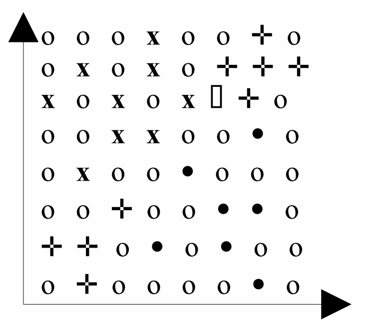

Fig. 1. Example of a 2D observation space with elements representing three similarity classes..

2^{nd} in columns. Starting from the lowest left element we can observe that the classified elements occur in the following order:

▬ • ▬ ▬ • •▬ • • x • x x • x x x ? ▬ x x ▬ ▬ ▬ x ▬

where the position of the unknown element **?** also has been indicated. It should be recognized as such that minimizes the number of homogenous sub-series in the given sequence of elements. Till the element **?** is neglected the number of sub-series is 16. Including **?** into the sequence leads to the following three possible solutions (the corresponding numbers of sub-series are given in the brackets):

▬ • ▬ ▬ • • ▬ • • x • x x • x x x | x | ▬ x x ▬ ▬ ▬ x ▬ (16)
▬ • ▬ ▬ • • ▬ • • x • x x • x x x | • | ▬ x x ▬ ▬ ▬ x ▬ (17)
▬ • ▬ ▬ • • ▬ • • x • x x • x x x | ▬ | ▬ x x ▬ ▬ ▬ x ▬ (16)

Therefore, it can be concluded that in the given case **?** can be recognized either as **x** or as ▬. Let us stress it out that:

1. The method can be applied to any non-homogenous (i.e. representing data of various formal nature) multi-dimensional observation space;
2. it is independent on observation space scaling;
3. it suits to any finite number of similarity classes, their merging into higher-order classes or splitting into similarity subclasses can be easily reached;
4. the method can be easily implemented on computers;
5. storage of the learning sets' elements in computer memory is not needed, instead of this the rules of observation space linear ordering should be stored.

However, it arises a problem of an adequate to the recognition problem linear ordering of the observation space. In the next sections of this paper the problem will be considered in a more detailed form.

2 Impact of linear ordering on pattern recognition efficiency

In this section it will be illustrated the influence of linear ordering introduced to observation space on the effectiveness of pattern recognition. For this purpose the 2D observation space and the training sets shown in Fig. 1 will be considered. In Fig. 2 several examples of linear ordering are shown: a/ lexicographical, b/ reversible sequential, c/ diagonal and d/ spiral. By symmetrical reflection with respect to a vertical axis and rotation through the angle -90^o the orderings a/, b/ and c/ can be transformed into the alternative ones based on different direction of observation space scanning; they are denoted, respectively, by a'/, b'/ and c'/. For the formerly given training sets S_1, S_2 and S_3 and the defined linear orderings the corresponding sequences of training elements and the numbers of homogenous sub-series can be calculated:

a/ ▬ • ▬ ▬ • • ▬ • • x • x x • x x x ▬ x x ▬ ▬ ▬ x ▬ (16)
a'/ ▬ x ▬ ▬ x x ▬ x x • x x x • x • • ▬ • • • ▬ ▬ ▬ ▬ (14)
b/ ▬ x x x ▬ ▬ ▬ x x x x x • • x ▬ • • • • • ▬ ▬ ▬ ▬ (9)
b'/ ▬ • • • ▬ ▬ ▬ • • • x x x • ▬ x x x x x ▬ ▬ ▬ ▬ x (10)
c/ • • • • • • ▬ ▬ • ▬ ▬ ▬ ▬ ▬ x x ▬ ▬ x x x x x x x (7)
c'/ ▬ ▬ ▬ x ▬ • x x • • • • x x x • x x x • ▬ ▬ ▬ ▬ ▬ (11)
d/ ▬ x x ▬ ▬ • ▬ ▬ x x x ▬ ▬ ▬ • • • • ▬ x x x • x • (13)

It can be observed that in the given case ordering c/ leading to the lowest number (7) of sub-series is the best one; ordering b/ (9 sub-series) also belongs to satisfactory ones. However, neither c/ nor b/ is an optimal ordering, the minimum possible number of sub-series being equal 3—the number of distinguished similarity classes. This minimum should be taken into account as the aim of observation space linear ordering optimization. The methods of linear ordering shown in Fig. 2 a-d can be called regular, as defined by simple geometrical rules. They also can easily be extended on multi-dimensional observation spaces. Some other (a little more sophisticated) regular ordering methods preferring local space scanning and based on a general Hilbert concept of curves filling compact geometrical areas, have been presented in [5].

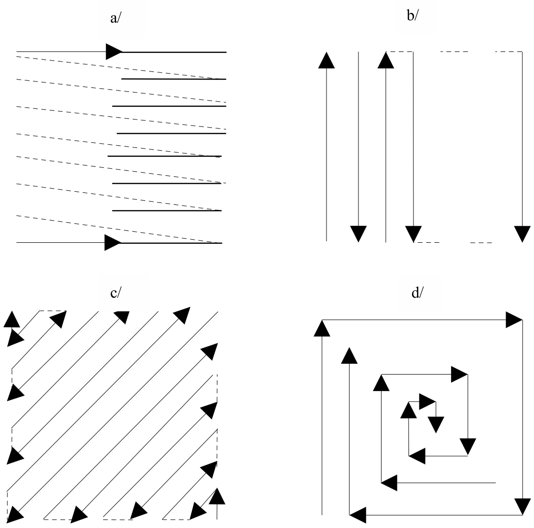

Fig. 2. Selected types of discrete 2D observation space linear orderings: a/ lexicographical (w. r. to columns and rows), b/ reversible sequential, c/ diagonal, d/ spiral.

However, simple orderings usually do not satisfy the optimality criteria and this is why it arises the problem of linear ordering optimization. A heuristic solution of this problem based on hyper-cubes permutations approach has been proposed in [6]; in this paper a more general approach to the problem is presented. In particular, choosing the best linear ordering is considered as an optimization problem that should take into account both the low implementation complexity costs and high pattern recognition effectiveness requirements.

3 Linear ordering optimization

It will be considered a discrete multi-dimensional observation space defined as a Cartesian product of a finite number of discrete sub-spaces:

$$\mathbf{D} = D_1 \times D_2 \times \ldots \times D_n \tag{1}$$

the sub-spaces D_ν, $\nu = 1, 2, \ldots, n$, as representing various features of objects under observation, may have different formal nature and are assumed to be linearly ordered independently each on each other one. In addition, taking into account computer implementation of pattern recognition systems it is assumed that each feature represented by a discrete scale D_ν takes values from a finite interval; by simple transformation this interval can be represented in a standard form of a sequence of integers $\Delta_\nu = [1, 2, \ldots, m_\nu]$. Therefore, in practice the real observation space can be reduced to a discrete hyper-cube:

$$\Delta = \Delta_1 \times \Delta_2 \times \ldots \times \Delta_n \tag{2}$$

The elements of each hyper-cube of this type can be ordered linearly in $M!$ ways where $M = (m_1 \cdot m_2 \cdot \ldots \cdot m_n$ is the number of elements of Δ. A lexicographical ordering of Δ based on a fixed order of its components will be distinguished as a one that can be easily technically realized. However, permutation of the components of $\varDelta$ results in changing the lexical order excepting the situation when the order is only reversed. Therefore, on the basis of the given family $F = \{\varDelta_\nu\}$ of n discrete sets the number $c = 1/2(n!)$ of lexicographical orders, significantly different from the pattern recognition point of view, can be established.

Let us denote by Q_M a sequence (finite linearly ordered set) of M elements. It is assumed that the pattern recognition task relies on assigning to any new-observed element $\mathbf{x}$, $\mathbf{x} \in \varDelta$, a similarity-class-index k, $k = 1, 2, \ldots, K$, K being a natural number > 1, so as to minimize, in a long series of experiments, the false recognition probability. In a supervised learning pattern recognition system it is assumed that to a certain sub-set $S \subset \varDelta$, called a training subset, pattern-indices have been a priori assigned. Therefore, to the elements of Q_M there can be assigned indices from an extended set $[0, 1, 2, \ldots, K]$, 0 being assigned to the elements not belonging to S. The indices assigned to Q_M form also a sequence $V(Q_M)$ which can be displayed in the following form:

$$V(Q_M) = [\nu_1, \nu_2, \nu_3, \ldots, \nu_M] \tag{3}$$

On the other hand, $V(Q_M)$ can be represented as a sequence of series of elements:

$$V(Q_M) = [\sigma_1, \sigma_2, \sigma_3, \ldots, \sigma_r] \tag{4}$$

where σ_ρ, $\rho = 1, 2, ..., r$, is a series of elements of a fixed value delimited on the left and on the right hand by other-value elements or by the ends of the sequence.

Theorem 1

If $M \geq K + 1$ then:

a) the minimum value of r is $r_{min} = K + 1$;

b) the maximum value of r is

$$r_{max} = N_1 \cdot (K + 1) + (N_2 - N_1) \cdot K + \ldots + (N_K - N_{K-1}) \cdot 2 + 1 \tag{5}$$

where $N_1,\ N_2, \ldots, N_{K+1}$ are the numbers of elements of $V(Q_M)$ of a given value, taken in a non-decreasing order, $N_1 \leq N_2 \leq \ldots \leq N_{K+1}$.

Proof

Part a) is evident because the elements of $V(Q_M)$ can be linearly ordered so that the elements tagged by "0" are taken first, the ones tagged by "1" are taken as the next ones, etc.

For proving part b) it will be, at the beginning, assumed that a series of strong inequalities $N_1 < N_2 < \ldots < N_{K+1}$ is satisfied. Then, at the 1st step, we can form N_1 sub-sequences consisting of elements tagged differently by $K + 1$ indices. Each sub-sequence of this type thus represents $K + 1$ one-element series. This justifies the first term of the right side of (5). After this operation there remain only elements tagged by K indices and the less numerous uniformly tagged subset contains only $N_2 - N_1$ elements. The former way of reasoning can be thus used again to the, so reduced sequence of elements, which leads to the second term of (5). Such operations can be repeated up to the moment when there remain only elements tagged by 2 different indices; their number $(N_K - N_{K-1}) + (N_{K+1} - N_K)$ makes us able to form $(N_K - N_{K-1})$ two-element sub-sequences of differently tagged elements. At last, there remain $(N_{K+1} - N_K)$ uniformly tagged elements which make us able to form only one series represented by the last term of (5).

In order to complete the proof let us assume that among the series of inequalities some weak inequalities occur; let it be, for example $N_1 = N_2$. Then after forming N_1 sub-sequences consisting of elements differently tagged by $K+1$ indices, in the next step there can be formed $N_3 - N_2$ sub-sequences consisting of K-1 differently tagged elements. This means that second term in the right side of (5) vanishes and, as a consequence, the number r_{max} is reduced •

For effective pattern recognition observation space should be linearly ordered so as to minimize the number r of series constituting the sequence $V(Q_M)$. This goal can be reached by a series of transformations of the initial lexicographical order. Each transformation is, in fact, a permutation of the elements of $V(Q_M)$. It is well known that each multi-element permutation can be realized by a sequence of simple permutations of pairs of the sequence elements. However, such linear order optimization as an NP-difficult numerical task, would be, in general, extremely time-consuming. That is why we are interested rather in looking for sub-optimal solutions of the problem.

We shall use, in general, a notion t for a transformation (simple or composed permutation) of the elements of (Q_M) (as well as of $V(Q_M)$). A sequence $t_a t_b \ldots t_f\{\ \}$ will denote the result of a consecutive application to Q_M the transformations: $t_f, \ldots, t_b$ and t_a.

The following types of transformations will be considered:

1. *reversion*: $t^r(p,q)$, $1 \le p < q \le M$, consists in taking from Q_M its segment (a compact sub-sequence) starting from the p-th element and ended by the q-th element, reversing their order and inserting them into the same place in Q_M;
2. *shifting*: $t^s(p,q,y)$, $1 \le p,q,z \le M$, $p < q$, $p \ne y$, consisting in taking from Q_M its segment starting from the p-th element and ended by the q-th element and shifting it in Q_M to a position starting from y;
3. *segmental permutation*: $t^p(p,q,y,z)$, $1 \le p,q,y,z \le M$, $p < q$, $y < z$, $p \ne y$, consisting in taking from Q_M its segment starting from the p-th element and ended by the q-th element, as well as taking a segment starting from the r-th element and ended by the s-th element, and mutual exchanging their position in Q_M.

It will be shown that the above-described transformations can be used to the improvement of an initial linear order of Δ.

Theorem 2

Let r be the number of series in $V(Q_M)$. Then the number r' of series in $t^r(p,q)\{V(Q_M)\}$ satisfies the inequality:

$$r - 2 \le r' \le r + 2 \tag{6}$$

Proof

Let us consider a part of the sequence $V(Q_M)$ containing the interval [p,q]. We shall denote by a,b,c,d, the values of this sequence delimiting the interval [p,q]:

$$\begin{array}{lllllll} Position: \ldots & p-1 & p & \ldots & q & q+1 & \ldots \\ Value: \quad \ldots & a & b & \ldots & c & d & \ldots \end{array}$$

where $a,b,c,d \in [0,1,2,\ldots,K]$. The following situations will be taken into account:

a) $b = c$ meaning that it is

$$\begin{array}{lllllll} Position: \ldots & p-1 & p & \ldots & q & q+1 & \ldots \\ Value: \quad \ldots & a & b & \ldots & b & d & \ldots \end{array}$$

Then a reversion of the segment $[p,q]$, independently on the values a and d, does not change the number of series, $r' = r$;

b) $a = d$ which, for similar reasons, leads to $r' = r$;

c) a ,b, c, d are all different, then reversion of the segment [p,q] also does not change the number of series, i.e. $r' = r$;

d) $b \ne c$, $a \ne d$, $a = b$, $c \ne d$ (or $a \ne b$, $c = d$) meaning that a situation:

$$\begin{array}{lllllll} Position: \ldots & p-1 & p & \ldots & q & q+1 & \ldots \\ Value: \quad \ldots & a & a & \ldots & c & d & \ldots \end{array}$$

will be changed into:

$$\begin{array}{lllllll} Position: \ldots & p-1 & p & \ldots & q & q+1 & \ldots \\ Value: \quad \ldots & a & c & \ldots & a & d & \ldots \end{array}$$

and, thus, the number of series will be increased, $r' = r + 1$;

e) $b \ne c$, $a \ne d$, $a = c$, $b \ne d$, (or $a \ne c$, $b = d$), the situation being exactly a reversion of this described in d) and, thus the number of series will be decreased, $r' = r - 1$;

f) $b \ne c$, $a \ne d$, $a = b, c = d$, i.e.:

$$\begin{array}{lllllll} Position: \ldots & p-1 & p & \ldots & q & q+1 & \ldots \\ Value: \quad \ldots & a & a & \ldots & d & d & \ldots \end{array}$$

In this case reversion of the segment [p,q] destroys two series and as a consequence the number of series is increased, $r' = r + 2$;

g) $b \ne c$, $a \ne d$, $a = c$, $b = d$, i.e.:

$$\begin{array}{lllllll} Position: \ldots & p-1 & p & \ldots & q & q+1 & \ldots \\ Value: \quad \ldots & a & d & \ldots & a & d & \ldots \end{array}$$

The situation is exactly a reversion of this described in f), therefore the number of series will be decreased, $r' = r - 2$.

In the case when $p = 1$ or $q = M$ the impact of the transformation $t^r(p, q)$ on the number of series is not greater than in the above-analyzed cases. The above-described situations thus complete the proof •

Corollary 1

Situation described in g) of the Proof of Theorem 1 recommend the most effective reversion as a transformation improving the linear order in Δ.

Theorem 3

Let r be the number of series in $V(Q_M)$. Then the number r' of series in $t^s(p, q, y)\{V(Q_M)\}$ satisfies the inequality:

$$r - 3 \leq r' \leq r + 3 \tag{7}$$

Proof

Like in the Proof of Theorem 2 it will be analyzed a part of the sequence $V(Q_M)$ containing the segment [p,q] and the element y as well as their close environments:

$$\begin{array}{llllllllll} Position: & \ldots & p-1 & p & \ldots & q & q+1 & \ldots & y-1 & y \\ Value: & \ldots & a & b & \ldots & c & d & \ldots & e & f \end{array}$$

where $a, b, c, d, e, f \in [0, 1, 2, , K]$. Shifting of the segment [p,q] leads to the following situation:

$$\begin{array}{llllllllll} Position: & \ldots & p-1 & p & \ldots & y-q+p-2 & y-q+p-1 & \ldots & y-1 & y \\ Value: & \ldots & a & d & \ldots & e & b & \ldots & c & f \end{array}$$

An analysis similar to this shown in the Proof of Theorem 2 leads to the conclusion that the following situations:

a) all values a, b, c, d, e, f are different,

b) all values a, b, c, d, e, f are equal,

c) $a = e$ and $d = f$ keep the number of series after shifting transformation unchanged, $r' = r$, because shifting of the segment [p,q] does not change its close environment.

In other cases the relationships within the pairs of values (a, b), (c, d), (e, f), (a, d), (e, b) and (c, f), as having a direct influence on the structure of series before and after the transformation, are critical for changing their number. From this point of view the following extreme situations can be distinguished:

d) if $a = b$, $c = d$, $e = f$, $a \neq d$, $e \neq b$ and $c \neq f$ then three series are split by insertion of segments of elements of other value and, as a consequence, we obtain $r' = r + 3$;

e) if $a \neq b$, $c \neq d$, $e \neq f$, $a = d$, $e = b$ and $c = f$ then three pairs of series are merged and, as a consequence, we obtain $r' = r - 3$.

In the case when $p = 1$ or $y = M$ the impact of the transformation $t^s(p, q, y)$ on the number of series is not greater than in the above-analyzed cases. The above-described situations thus complete the proof •

Corollary 2

Situation described in e) of the Proof of Theorem 2 recommends the most effective shifting as a transformation improving the linear order in Δ .

Theorem 4

Let r be the number of series in $V(Q_M)$. Then the number r' of series in $t^p(p, q, y, z)\{V(Q_M)\}$satisfies the inequality:

$$r - 4 \leq r' \leq r + 4 \tag{8}$$

Proof

The proof is similar to this shown in Theorem 3, therefore, it will be omitted.

Corollary 3

Situation described in e) of the Proof of Theorem 3 recommends the most effective segmental permutation as a transformation improving the linear order in Δ.

The Corollaries 1–3 give recommendations concerning choosing sequences of transformations of an initial linear order in observation space. However, the criterion of effectiveness of transformations was only on the number r of series. Therefore, it did not take into account the cost of calculations connected with using serial statistical tests based on the given linear order. The cost is the higher the larger is the distance between the initial lexicographical order and the linear order obtained as a result of a series of transformations. The distance between two linear orders of sequences consisting of the same elements is here understood as the minimum number of pair-wise permutations transforming one order into the other one. On the other hand, from the Theorems 1–3 no recommendations follow about choosing the position and lengths of reversed, shifted or permuted segments. However, it seems reasonable to operate, first of all, on the segments delimited by the initial lexicographical order.

4 Conclusions

Linear ordering of observation space is a new paradigm of pattern recognition based on non-parametric, serial statistical tests. Effectiveness of any test of this type depends on choosing a linear order appropriate to the geometrical form of similarity classes, which are a priori unknown. Therefore, it arises the problem of linear order optimization according to the available learning data-subsets. In this paper it was shown how to improve the initial linear order (usually lexicographical) by a sequence of transformations: reversing the order in selected intervals, shifting the intervals or permutation of selected pairs of intervals on the sequence of linearly ordered data. The considerations have rather a theoretical character, however, they indicate the way to the construction of the corresponding algorithms which will be used in pattern recognition systems based on serial statistical tests.

Acknowledgement

The Author would like to express his thanks to Mrs Diana Wierzbicka for her valuable help in editorial preparing the manuscript of this work.

References

1. Patrick E.: *Fundamentals of Pattern Recognition,* Prentice-Hall, Inc., New York, 1972.
2. Duda R., Hart P.: *Pattern Classification and Scene Analysis,* John Wiley & Sons, New York-London-Sydney, 1973.
3. Tou J., Gonzales R.: *Pattern Recognition Principles. Addison-Wesley Publishing Co,* London-Amsterdam-Sydney, 1974.
4. Kulikowski J. L., Wierzbicka D.: *Texture Analysis Based on Application of Non-Parametric Serial Statistical Tests,* Biocybernetics and Biomedical Eng., Vol. **24**, No 2, 2004, pp. 27–39.
5. Kulikowski J. L., Wierzbicka D.: *Choosing Serial Tests for Discrimination of Textures in Biomedical Images,* Biocybernetics and Biomedical Eng., Vol. **25**, No 3, 2005, pp. 65–77.
6. Kulikowski J. L., Przytulska M., Wierzbicka D.: *A Method of Supervised Discrimination of Textures Based on Serial Statistical Tests,* Computer Recognition Systems (M. Kurzynski et al. Eds.), Proc. of the 4th Int. Conf. on Computer Recognition Systems CORES'05. Springer-Verlag, Berlin, 2005, pp. 235–242.