RECOGNITION RULE FOR NORMALLY DISTRIBUTED VECTORS AFTER SECONDARY ORTHOGONAL TRANSFORMATIONS

Dimitar St. Tyanev

Abstract: *On the basis of a critical analysis of the orthogonal transformations used in the problems of pattern recognition, a new method of approach is suggested, which is meant for the optimization of the orthogonal feature spaces. This approach consists in accomplishment of consecutive secondary orthogonal transformations by manipulating the parametres of the eigen basis systems to the optimal point of view. The latter is determined by the use of the term secondary scalar feature. In connection with the introduced presentation of classes, a classification rule is formulated.*

Key words: Orthogonal feature space; Orthogonal transformations; Pattern recognition problems.

1. Problems of the traditional approach

The recognition of a given pattern **x** (in its abstract sense) is a process which results in the degree of its similarity in relation to a certain multitude of **p**-possible for it Ω classes. Here that means classes of normally distributed in **k-**dimensional space vectors, represented by their own distribution laws:

$$\textbf{(1)} \qquad N(\mathbf{C}^{(\Omega)}, \mathbf{m}^{(\Omega)}), \quad \Omega = \overline{1,p} \ ,$$

the parametres of which are the co-variational matrices **C** and the vectors of the mathematical expectancy **m**.

Their similarity or also their counterpart is assessed by the means of different formal approaches in the sense of various mathematical criteria. However, the strength of those criteria is directly dependent on the informative abilities of the pattern-descriptive formal features. That's why there has always been an aspiration towards transitions into such feature spaces which are optimum in the sense of informativity or, in other words, in these spaces the structural features of a given class are better expressed. One of the most frequently applied data transformations from the primary feature space are the orthogonal ones. It's known [1], that in conformity with the expressed concept, the optimum orthogonal space is the one of the main components. The pattern projection in that space is performed through the linear operator of the $\mathbf{y} = \mathbf{A}^T.\mathbf{x}$ kind, where the matrix $\mathbf{A}^T$ is the orthonormed vector basis of the new space. This transformation gives the opportunity that every separate class to be represented by its own feature system. Simultaneously with the receiving of the vector basis, the problem for choice of the new features is also solved. Practically this solution is achieved by choosing only a part of the main components thus forming a sub-space $\mathbf{A}_0$, called the Kharunen-Loeve space. Then for recognition of a given pattern projected in the Kharunen-Loeve space, is necessary that an evaluation for that is received, how precisely it is represented in the space of every separate class. The pattern belongs to the class in which space it is depicted most accurately. The formal mathematical apparatus for receiving of such evaluations is known [1÷6, etc.].

The informative properties of the space of the main components are directly expressed through the eigenvalues $\lambda_i, i = \overline{1,k}$ of its determining co-variational matrix **C**. As these values are real numbers and each one of them is connected with an exactly defined basis vector (feature), that gives a chance for the informative part of every single feature to be easily assessed.

The immediate construction of the linear operator through which the data is being depicted in the new feature space, requires that an orthogonal decomposition of the co-variational matrix of each class is performed separately:

$$\textbf{(2)} \qquad \mathbf{C}^{(\Omega)} = \mathbf{A}^{(\Omega)}.\mathbf{D}^{(\Omega)}.\left(\mathbf{A}^{(\Omega)}\right)^T, \ \Omega = \overline{1,p}$$

where p is the number of pattern-known classes. If the matrix elements

(3) $$\mathbf{D}^{(\Omega)} = \mathrm{diag}\left(\lambda_i^{(\Omega)}\right), \quad i = \overline{1,k} ,$$

are arranged in a descending line, then the orthonormed matrix $\mathbf{A}^{(\Omega)}(k,k)$ may be split into two blocks:

(4) $$\mathbf{A}^{(\Omega)}(k,k) = \left[\ \mathbf{A}_0^{(\Omega)}(k,n) \ \vdots\ \mathbf{A}_1^{(\Omega)}(k,k\text{-}n)\ \right]$$

in conformity with the estimative splitting of the **D** matrix into two parts:

(5) $$\mathbf{D}^{(\Omega)} = \begin{bmatrix} \mathbf{D}1^{(\Omega)}(n,n) & \vdots & \mathbf{0} \\ \cdots & \vdots & \cdots \\ \mathbf{0} & \vdots & \mathbf{D}2^{(\Omega)}(k-n,k-n) \end{bmatrix}$$

at which the $\mathbf{D}2^{(\Omega)}$ matrix involves the assessed as insignificant (k-n) in number eigenvalues. The number $n^{(\Omega)}$ is individual for every single class. It shows the dimension of its own space assumed as significant for the given class.

The setting of dividing boundary between the elements of the matrix **D** is based on the a priori evaluation of the tolerance error with which the transformed data could be used. The loss of accuracy which could be accepted as tolerable depends on the object requirements and the possibilities of the engineering application.

The determination of the dividing boundary could be facilitated if the elements of the matrix **D** in their first part are predominant. Thus they would naturally be divided into significant and insignificant. However, practically, under the circumstances of most of the real objects studied by us [10, 11, 12, 13], etc., the obtained eigenvalues are close to each other and have weak fading, i.e., the primary feature spaces do not keep a pronounced feature grouping which to promote for the splitting naturally in the sense of (5). Such situations can be explained only by the indirect character of the primary features regarding the physical nature of the real object. In such cases the choice as to which basis vector to be ignored in virtue of the relative comparison of an eigenvalue with another one is made highly difficult and inobjective. The choice is made even more difficult when the eigenvalues are divisible which means that the eigen vectors to which they correspond are equivalent with respect to the criterion of choice. This case requires a special attention as it leads to more than one solution for the factorization (2) [10]. The problem for finding an eigen space of the main components is naturally connected with the problem for decreasing of its dimension, and it is much easier to solve than in the primary feature space, where for evaluation of the informative capacity of feature are applied various difficult methods [11]. At the same time the selection of features according to the criteria connected with the inner-class and inter-class distance requires a simultaneous pattern processing by all of the classes.

These peculiarities, as well as some difficulties in the calculation process with the realization of decomposition (2), forces some research scientists towards use of other orthogonal transformations like those of Fourier, of Walsh or of Haar. It is also known that for separate classes of probability processes these transformations are optimum in the sense of Kharunen-Loeve transformation. Despite that the receiving of informative features with the help of a single orthogonal transformation in most of the cases in our opinion is having a difficulty or, is not quite successful and useful. The pointed out shortcomings for data presentation in the individual space of the main components, as well as the difficulties connected with its interpretation, suggests to us the conclusion that this space is not yet the most appropriate.

Though the structural characteristics of classes are clearer remarkable in the space of the main components than in the primary feature space, we consider that it is still not the utmost possibilities of this approach. We are definitely justified to say that because of the possibility when there is a suitable interpretation of a given orthogonal basis system and by the means of appropriate control of its parametres [7], [8], to be formed another basis system in which the structural characteristics of a given class could be simpler and clearer. In other words, there exists the conviction that a new viewpoint could be found for observing the distribution structure. Thus for each class different in character and interpretation could be found eigen parametrically-controlled bases. Below a method of approach for manipulation of basis systems is stated.

2. General approach to formation of secondary scalar features

Let by the means of (1), a number of p classes are known by their k-dimensional distributions in primary feature space which is common to all classes. Every co-variational matrix has the decomposition (2) and could be used for construction of eigen k-dimensional vector basis $\mathbf{A}^{(\Omega)}$ of the main components space. Through the transformation:

(6) $$\mathbf{y}_j^{(\Omega)} = \left(\mathbf{A}^{(\Omega)}\right)^T . \mathbf{x}_j^{(\Omega)}$$

every j-th pattern of the observed in this primary feature space statistics is projected in the new feature space which is a space only to the given class Ω. Thus it becomes possible that every separate class be represented by the pair

(7) $$\left\{\mathbf{A}^{(\Omega)}, \boldsymbol{\mu}^{(\Omega)}\right\}, \quad \Omega = \overline{1,p} \ ,$$

i.e., by its vector basis and the vector of the mathematical expectancy μ. The arranged pair (7) is in condition to present the given class, as it, together with the calculated eigenvalues, expresses its statistical structure and is in full conformity with the law (1). For the distribution of a given class in this space it could be said that it is oriented towards the directions of the co-ordinate axes and is centred to the vector of mathematical expectancy μ. In this sense, the distribution is to be found in a general position in relation to the observing direction determined by the μ vector. The transformation (6) and the presentation (7) can be realized also in the space of Kharunen-Loeve, if there are suitable conditions for that.

As a result to the motives stated in item 1, regarding the possibility to be found the better viewpoint, here the following problem is formulated: to be found such a secondary orthogonal basis **B**, in which the vector of mathematical expectancy ν to be co-linear to its primary basis vector. This means that the main direction of the equally probable ellipsoids of distribution in this new space will be oriented in the direction of the observer, i.e., in the direction of the vector of mathematical expectancy ν. At that, its projection in this space would have only one coefficient other than zero.

In order to solve the assigned problem, the formalized presentation of a given class through (7) is put to a parametrical reproduction of the vector basis by applying of generalized matrix cores of the following kind:

$$\begin{bmatrix} e^{i.\varphi_1}.\cos\theta & -e^{i.\varphi_2}.\sin\theta \\ e^{-i.\varphi_2}.\sin\theta & e^{-i.\varphi_1}.\cos\theta \end{bmatrix} ,$$

where θ, φ_1, φ_2 are the real parametres which reflect the desired properties of the synthesized basis [4].

As we already pointed out, hereby we aim at realizing such a reproduction of the vector basis **A** or $\mathbf{A}_0$ at which every class could be represented by the new pair

(8) $$\left\{\mathbf{B}^{(\Omega)}, s_1^{(\Omega)}\right\} , \quad \Omega = \overline{1,p} \ ,$$

where $s_1^{(\Omega)} \neq 0$ is a scalar. The arranged pair (8) depicts by virtue our wish to “observe” the structure of the given class from the most appropriate viewpoint. We will call that pair a secondary scalar feature.

The solving of the assigned task is reached through the use of Householder’s transformation. In [9] shows that for every k-dimensional vector **y** there exist such consequences from matrices of field rotations

(9) $$\mathbf{T}_{l_1 q_1}, \mathbf{T}_{l_2 q_2}, \ \ldots \ , \mathbf{T}_{l_r q_r}, \qquad r \leq k-1 ,$$

at which some of their product **B** turns the vector **y** into a vector that is co-linear to another known vector **z**, possessing the same Euclidean norm.

In this sense, taking into account the above-mentioned task, without any loss of universality we can consider that the solution we are looking for, is $\mathbf{z} = s_1.\mathbf{e}_1$, where s_1 is a scalar and $\mathbf{e}_1 = [1,0,\ldots,0]^T$. Then the linear operator

$$\mathbf{B}.\mathbf{y} = \mathbf{z} = s_1.\mathbf{e}_1 \tag{10}$$

will be called by us a secondary orthogonal transformation. As this transformation is invariant in relation to the Euclidean norm, then the number s_1 shows exactly that norm. In other words

$$s_1 = (\mathbf{y}^T.\mathbf{y})^{1/2} \quad , \tag{11}$$

at which both values $s_{11} > 0$ and $s_{12} < 0$.

The algorithm for forming the matrix **B** is shown in Eq.[14].

3. Formulating the recognition rule

The transformation (10) which is to be applied on the observed patterns **x**, turning in accordance with the scheme $(\mathbf{x} \Rightarrow \mathbf{y} \Rightarrow \mathbf{z})$, requires formulating of a recognition rule. For that purpose here it is suggested a function for evaluation of accuracy with pattern presentation in the new feature space. The synthesis reasoning of that function goes as follows: as every separate class is going to have its own individual transformation form the kind (10), then the patterns belonging to a given class will be represented in it with a predominant first element. For all other vectors, the all elements left will prove to be significant. By force of that reasoning (which is also the reasoning of the synthesized space from the **B** type), the degree of accuracy with which the $\mathbf{y}^{(\Omega)}$ vector changes into the secondary space $\mathbf{B}^{(\Psi)}$ of the class Ψ, $(\Omega \neq \Psi)$ will be assessed by us through the function G, determined as follows:

$$G^{(\Omega)} = \sum_{i=2}^{k} \left(z_i^{(\Omega)} \right)^2 , \quad \Omega = \overline{1,p} \tag{12}$$

We use the function G in order to make up the next recognition rule:

$$\textbf{\textit{if}} \quad J_{\Omega\Psi} = G^{(\Omega)}(\mathbf{x}) - G^{(\Psi)}(\mathbf{x}) < 0 \quad \textbf{\textit{for every}} \quad \Omega \neq \Psi, \quad \Omega, \Psi = \overline{1,p}$$
$$\textbf{\textit{then}} \quad \mathbf{x} \in \Omega$$

4. Final notes

In closing the exposition of the formal side of the above-treated task and its application, we are going to comment more freely on the obtained results. Essentially, a brief summary of recent researches of ours was exposed. Those researches are connected with the real application of various methods from the theory for pattern recognition, as well as the main conclusions which we drew after critically assessing them. It is, of course, natural that every summary will evoke new ideas, as the case goes with the suggested here approach to transition towards secondary orthogonal eigen basis systems. That approach gives the opportunity for choosing the most appropriate viewpoints in relation to the structure of every class. However, the introduced presentation of classes by the means of the secondary scalar feature (8) is not the only way for presenting them - it can be supplemented by a second and even a third scalar, and it can obtain the following kind:

$$\left\{ \mathbf{B}^{(\Omega)}, s_1, s_2, s_3 \right\} , \tag{13}$$

which means that the class structure obtained in the space **B** could be observed in a single two-dimensional or even three-dimensional sub-space, at that without loss of its synthesis reasoning. The scalars s_1, s_2, s_3 could be chosen in without restrictions in the algorithm for forming of the eigen secondary basis.

This method of approach releases us of the problem (5) for choosing the main components. It could even be claimed that it allows quite drastic solution of that problem, taking into consideration both presentations (8) or (13).

Literature:

[1]. Ватанабе С., *Разложение Карунена-Лоева и факторный анализ. Теория и приложения*. В книге “Автоматический анализ сложных изображений”, Москва, Издательство “Мир”, 1969.

[2]. Tou J.T., Gonzalez R.C., *Pattern recognition principles*, Addison-Wesley pub. company, London, 1974.

[3]. Дуда Р., Харт П., *Распознавание образов и анализ сцен*, Москва, Издательство “Мир", 1976.

[4]. Патрик Э., *Основы теории распознавания образов*, Москва, Издательство ‘Советское радио", 1980.

[5]. Браверман Э. М., Мучник И. Б., *Структурные методы обработки эмпирических данных*, Москва, Издательство “Наука", 1983.

[6]. Фомин Я. А., Тарловский Г. Р., *Статистическая теория распознавания образов*, Москва, Издательство “Радио и связь", 1986.

[7]. Fukunaga K., *Introduction to statistical pattern recognition*, 2-nd ed., Academic press, New York, 1990.

[8]. Дагман Э. Е., Кухарев Г. А., *Быстрые дискретные ортогональные преобразования*, Академия наук СССР, Издательство “Наука", 1983.

[9]. Солодовников А. И., *Синтез полных систем ортонормированных функций имеющих быстрого преобразования*, В книге “Вопросы теории систем автоматического управления”, Ленинградский университет, 1978.

[10]. Horn R.A., Johnson C.R., - *Matrix analysis*, Cembridge university Press, 1986.

[11]. Тянев Д. С., *Оценка на разделящите свойства на признаците при синтез на разпознаващи алгоритми*, Годишник на ТУ-Варна, 1986.

[12]. Tyanev D.S., Dobrev P.D., - *Two chanel 16-inputs A/D controller with DMA for IBM PC/AT*, 8-th International conference "Systems for automation of engineering and research", X.1994, Varna.

[13]. Недев А., Тенекеджиев К., *Техническа диагностика и разпознаване на образи*, Издателство ТУ – Варна, 1994.

[14]. Тянев Д. С., *Вторични ортогонални преобразования в задачите за разпознаване на образи*, ЮНС’96 “115 години ВВМУ “Н.Й. Вапцаров”, Варна, V.1996.