# Developing Methods and Algorithms for Forming of Informative Features' Space on the Base K-Types Uniform Criteria

**Fazilov Shavkat, Mamatov Narzillo, Niyozmatova Nilufar**

**Abstract**: *A new method for extracting the set of informative features on the base k-types uniform criteria is proposed in this paper.*

*Keywords: object, vector, operator, class, Fisher type uniform functional, uniform criteria, informative features.*

## I. INTRODUCTION.

The systems of pattern recognition are widely applied to the problems of controlling of objects. This is because modern complex technical objects often cannot be fully and correctly described by the mathematical equations. Controlling of objects that do not have strict mathematical description is very important problem.

Experience on creating of control systems on complex objects shows that it is necessary to use all the resources of classical theory of controlling of objects, but if it is not possible, adaptive, supervised and unsupervised learning systems, such as learning systems of pattern recognition, should be applied.

Developing of pattern recognition systems requires solving several complex problems, such as extracting of informative features from the set of features that provides maximum efficiency of the decisions obtained by control system on the base of solutions of the pattern recognition problem.

Choosing of informative features is a traditional pattern recognition problem. Many algorithms and methods have been developed for choosing of informative features with the help of 0-degree uniform criteria.

But in real-time problems, it is required to learn the problem of choosing of informative features with the help of k-degree uniform criteria.

And the problem of choosing of informative features with the help of k-degree uniform criteria is discussed in this paper.

## II. FORMULATION OF THE PROBLEM.

Following training sample is given:

$$x_{11}, x_{12}, \ldots, x_{1m_1} \in X_1;$$

$$x_{21}, x_{22}, \ldots, x_{2m_2} \in X_2;$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$x_{r1}, x_{r2}, \ldots, x_{rm_r} \in X_r,$$

Where $x_{pi} = \left(x_{pi}^1, x_{pi}^2, \ldots, x_{pi}^N\right)$ -objects of the class $X_p$, $p = \overline{1, r}$.

We define as: $\overline{x}_p = \dfrac{1}{m_p} \sum\limits_{i=1}^{m_p} x_{pi}$, $p = \overline{1, r}$.

Here and further, sum of the vectors is defined as a vector components of which are sums of matching components. Vector $\overline{x}_p$ is average object of the class $X_p$. We define average deviation for the class $X_p$ relatively to the subsystem of features given by vector $\lambda$:

$$S_p(\lambda) = \sqrt{\frac{1}{m_p} \sum\limits_{i=1}^{m_p} \left\| x_{pi} - \overline{x}_p \right\|_\lambda^2}.$$

Following functional is generalization of the Fisher's k-degree criteria for the multidimensional case, especially:

$$I(\lambda) = \frac{\sum\limits_{p,q=1}^{r} \left\| \overline{x}_p - \overline{x}_q \right\|_\lambda^2}{\prod\limits_{p=1}^{r} S_p^2(\lambda)} \quad (1)$$

For the given functional, subsystem of features is more informative, if value of the functional is bigger. As a result, functional can easily be transformed to the following:

$$I(\lambda) = \frac{(a, \lambda)}{\prod\limits_{i=1}^{r} (b^{(i)}, \lambda)} \quad (2)$$

Where $a = \left(a^1, a^2, \ldots, a^N\right)$, $b = \left(b^1, b^2, \ldots, b^N\right)$ -N dimensional vectors.

**Fazilov Shavkat[1], Mamatov Narzillo[2], Niyozmatova Nilufar[3]**
1,2Scientific and Innovation Center of Information and Communication Technologies at TUIT named after Al-Kharezmi, Tashkent, Uzbekistan
3Tashkent university of information technologies named after Muhammad al-Khwarizmi

$$a_j = \sum_{p,q=1}^{r}\left(\bar{x}_p^j - \bar{x}_q^j\right)^2; \; j=\overline{1,N}, \; b_j^{(i)} = \left[\frac{1}{m_p}\sum_{t=1}^{m_p}\left(\bar{x}_p^j - x_{pt}^j\right)^2\right]; \; j=\overline{1,N}, \; i=\overline{1,r}.$$

$a^j$, $b^j$ coefficients are independent from $\lambda$ and calculated beforehand. N operations are required for the calculation of $I(\lambda)$ functional for each $\lambda$.

Criterion for informativeness of the features given as a functional (2) is heuristic and based on the hypothesis of compactness which is fundamental hypothesis in pattern recognition.

For evaluation of informativeness of features heuristic criteria, especially criteria described as a functional (2) which can be easily maximized, are further used in this work.

Functional (2) is a generalized form of the Fisher type uniform functional with the degree $k$ $(k \leq 0)$.

We consider following searching problem of informative subsystem of features:

$$\begin{cases} I(\lambda) = \dfrac{(a,\lambda)}{\prod\limits_{i=1}^{r}\left(b^{(i)},\lambda\right)} \to max \\ \lambda \in \Lambda^l, \end{cases} \qquad (3)$$

where $\Lambda^l$ -set of all informative features $\ell$.
We define following to simplify calculations:

$$A = (a,\lambda), \; B^{(j)} = \left(b^{(j)},\lambda\right), j=\overline{1,r},$$

$$A_1 = (a,\mu), \; B_1^{(j)} = \left(b^{(j)},\mu\right), j=\overline{1,r}.$$

Then $I(\lambda) = \dfrac{A}{\prod\limits_{i}^{r} B^{(j)}}$, $I(\mu) = \dfrac{A_1}{\prod\limits_{i=1}^{r} B_1^{(j)}}$.

We define following gradient vector to solve (3)
$C = (c_1, c_2, ..., c_N)$:

$$c_i = \frac{a_i}{w^2} - I_\lambda \sum_{j=1}^{r} b_i^{(j)}, \; i=\overline{1,N}$$

Where $w = \max\limits_{j=1,r} \sum\limits_{i=1}^{n} b_i^{(j)}$;

$$I_\lambda = I(\lambda) = \frac{A}{\prod\limits_{i=1}^{r} B^{(j)}};$$

## III. SOLUTION OF THE PROBLEM

**Theorem 1.** If $\lambda$ and $\mu$ are two $\ell$ - informative vectors and $b_t^{(j)} > 0$, $t=\overline{1,N}$, $j=\overline{1,r}$, then $I(\lambda) < I(\mu)$ if and only if $(C,\mu) > 0$.

Proof.
Let $(C,\mu) > 0$. Then

$$(C,\mu) = \frac{(a,\mu)}{w^2} - I_\lambda \sum_{j=1}^{r}(b,\mu) > 0, \text{ or}$$

$$\frac{A_1}{w^{r-1}} > I_\lambda \sum_{j=1}^{r} B_1^{(j)} \qquad (4)$$

Right side of (4) we can express as:

$$v\sum_{j=1}^{r} B_1^{(j)} = I_\lambda w \sum_{j=1}^{r} \frac{B_1^{(j)}}{w} \qquad (5)$$

Considering $\sum\limits_{j=1}^{r} \dfrac{B_1^{(j)}}{w} < 1$, we have

$$I_\lambda w \sum_{j=1}^{r} \frac{B_1^{(j)}}{w} \geq I_\lambda w \frac{\prod\limits_{j=1}^{r} B_1^{(j)}}{w^r} = I_\lambda \cdot \frac{\prod\limits_{j=1}^{r} B_1^{(j)}}{w^{r-1}} \qquad (6)$$

From inequalities (6) and (4), we get following inequality:

$$\frac{A_1}{w^{r-1}} > I_\lambda \sum_{j=1}^{r} B_1^{(j)} \geq I_\lambda \cdot \frac{\prod\limits_{j=1}^{r} B_1^{(j)}}{w^{r-1}} \qquad (7)$$

We transform (7) as follows:

$$A_1 > I_\lambda \cdot \prod_{j=1}^{r} B_1^{(j)} \Rightarrow \frac{A_1}{\prod\limits_{j=1}^{r} B_1^{(j)}} > I_\lambda = \frac{A}{\prod\limits_{i=1}^{r} B^{(j)}} \Rightarrow I(\mu) > I(\lambda).$$

Thereby the theorem has been proven.
We define the operator $\mu: \Lambda^\ell \to \Lambda^\ell$ so that

$$(C,\mu(\lambda)) = \max_{\eta \in \Lambda^\ell}(C,\eta).$$

Operator $\mu$ has obviously constructive expression. If we put components of the vector $C$ in orderly manner so that to find a set of different pairs of $j_1, j_2, \ldots, j_N$ for which $c^{j_1} \geq c^{j_2} \geq \ldots \geq c^{j_N}$, then components of the vector $\mu(\lambda)$ could be defined as

$$\mu^{j_1}(\lambda) = 1, \mu^{j_2}(\lambda) = 1, \ldots, \mu^{j_\ell}(\lambda) = 1, \mu^{j_{\ell+1}}(\lambda) = 0, \mu^{j_{\ell+2}}(\lambda) = 0, \ldots, \mu^{j_N}(\lambda) = 0 \quad .$$

In other words, components of the vector $\mu(\lambda)$ corresponding to the first $\ell$-maximum components of the vector $C$ are equal to 1, remaining components are equal to 0.

Obviously, $\mu(\lambda)$ is $\ell$-informative vector, moreover:

$$(C, \mu(\lambda)) = \max\{(C, \eta)|\eta \in \Lambda^\ell\}. \quad (8)$$

**Property 1.** For the any $\lambda (\lambda \in \Lambda^l)$ $(C, \mu) > 0$ is true.

Proof. From (8)

$$(C,\mu) > (C,\lambda) = (C,\mu) = \frac{(a,\lambda)}{w^{r-1}} - I_\lambda \sum_{j=1}^{r}(b,\lambda) = \frac{A}{w^{r-1}} - I_\lambda w \sum_{j=1}^{r} \frac{B^{(j)}}{w} \geq \frac{A}{w^{r-1}} - I_\lambda \frac{\prod_{j=1}^{r} B^{(j)}}{w^{r-1}} = 0$$

Property has been proven.

Following main conclusion could be obtained from theorem 1 and property 1

$$I(\lambda) \leq I(\mu(\lambda)), \lambda \in \Lambda^\ell.$$

**Theorem 2.** If $I(\lambda) = I(\mu(\lambda))$, then $I(\lambda) = \max\{I(\eta)|\eta \in \Lambda^\ell\}$.

Proof. $(c, \mu(\lambda)) = 0$ could be obtained from theorem 1. Thereby, according to (8), we have $(C, \mu(\lambda)) = 0 = \max\{(C, \eta)|\eta \in \Lambda^\ell\}$ or $(C, \eta) \leq 0$ for any $\eta \in \Lambda^\ell$.

According to the theorem 1, this means that $I(\eta) \leq I(\lambda)$ for any $\eta \in \Lambda^\ell$, $I(\lambda) = \max\{I(\eta)|\eta \in \Lambda^\ell\}$.

Theorem has been proven.

It should be noticed that theorem 2 guarantees optimality of obtained result, that is the value of the functional $I(\lambda)$ for the obtained result $\lambda$ reaches to its maximum within the set $\Lambda^\ell$.

Proposed method for maximization of the functional (2), realized as an iterative process, is based on the theorems 1 and 2. As a first step, any $\ell$-informative vector $\lambda$, for example $\lambda = \left( \overbrace{1,1,\ldots,1}^{\ell}, 0,0,\ldots,0 \right)$, is chosen.

Further, a new vector $\lambda$ is defined from previous one with the help of the operator $\mu(\lambda)$ in each iteration so that simple assignment $\lambda = \mu(\lambda)$ is used.

Iterative process is continued until the functional $I(\lambda)$ continues to grow. If it stops to grow, that is $I(\lambda) = I(\mu(\lambda))$, then $\lambda$ is optimal solution. As experiments show, this result is obtained in steps 3 or 4.

## IV. CONCLUSION

As a conclusion, it should be noticed that obtained results are suitable for all criteria of informativity of features given by functionals that could be described by (2).

## REFERENCES

1. Fazilov Sh., Mamatov N.S., Method of "Deltas" for building of informative features' space. Uzbek Journal "The Problem of Informatics and Energetic". 2005, #6, pp.11-16.
2. Mamatov N.S. Working out the methods and algorithms for forming of informative features space with the help of k-types criteria., Uzbek Journal "The Problem of Informatics and Energetic". 2006, #2-3, pp.14-19.
3. Fazilov Sh., Mamatov N.S., Gradient method for forming a space of informative features based on uniform criteria with positive degree. Cybernetics and Informatics: reports of the II International Conference, Bishkek, 2007. Pp 8-9
4. Mamatov N.S., Working out the methods and algorithms for forming of informative features space with the help of k-types criteria. Reports of the Academy of Sciences of Republic of Uzbekistan.-2007.-№4. pp. 50-53.
5. Fazilov Sh., Mamatov N.S., Gradient method for forming a space of informative features based on uniform criteria with positive degree. Reports of the Academy of Sciences of Republic of Uzbekistan.-2008.-№2. pp