УДК 621. 372:621.314

DETERMINATION OF PRIORITY OF PRIMARY SIGNS AND ITS CORRECTION FOR CONSTRUCTION OF OBJECTS RECOGNITION SYSTEMS

V.M. ZAYATS, G.YA. SHOKYRA

The paper suggests an approach to determining the priority of primary signs when constructing objects recognition systems based on the method of successive approximations. The method of correcting the established priorities by adding new signs to the selected base is developed. The results obtained have been confirmed by making computer experiments when modelling elements of recognition systems with deterministic signs.

Keywords: recognition system, primary signs, working vocabulary, priority, correction method.

INTRODUCTION

At creation of the systems of recognition of objects and their reliable authentication a system approach is needed. Its essence lays in forming of primary signs of the object of recognition, establishment of their priority, reliable determination and realization of reliable criteria of authentication and recognition of objects.

The first researches in industry of recognition in our country were conducted by A.A. Kharkevich [1] – one of the founders of information theory. A considerable contribution to development of theory of recognition was done by V.M. Glushkov, V.S. Mikhalevich, O.G. Ivakhnenko, Yu. I. Zhuravlev, Yu. Z. Cipkin, V. I. Vasil'ev. Among the foreign scientists it is needed to remember a work Ф. Rosenblatt, which in 1957 offered a machine which studied to recognize the offenses and was named a perceptron. This was the simplest model of activity of human brain. The first works in the field of recognition of patterns were devoted to a theory and practice of construction of reading automats (under appearance a sign, image, letter or number, was understood). A mathematical machine for the decision tasks of recognition from the moment of their origin was a theory of statistical decisions [2].

The results of the theory of statistical decisions became a base for the construction of algorithms recognitions, which provided attributing of object to his class on the basis of a posteriori experimental information – signs, which characterize an object and a priori information which describe the classes of objects. Later a mathematical machine broadened due to the use of methods of algebra, logic and some sections of the applied mathematics, information, mathematical programming and analysis of the systems theory [3-8].

Despite to that methods and algorithms of recognition become the inalienable constituent of such applied industries of natural history, as medical and technical diagnostics, authentication of difficult hesitating processes, meteorological prognostication and geological secret service, ecological monitoring, in literature – both to domestic and in foreign – approach of the systems to the tasks of recognition absents so far.

Today, as well as 50 years ago, the problem of recognition equates with the construction of optimum algorithms of recognition and research of terms which allow realizing such algorithm. The researches of theorists are oriented on a decision though of important, but partial tasks. To such tasks in the first turn it is needed to take the tasks of reliable recognition, essence of which is taken to the division of space of signs the language of which is describe objects or processes of recognition, on areas which answer the classes of these objects, that to the choice of the best limits (rules) of division of classes. But the decision of these tasks is possible when the classes of objects and signs the language of which is the recognizable objects are described and their classes are known a priori. However, the developer of the system of recognition, as a rule, does not own this information. Even in the simplest cases of recognition of letters of alphabet, finger-prints, words of language, special points of functions (where a question is not about classes), them informing signs and apparatus for their determination are not known is s the article of untraditional researches.

Why then to the tasks of description of classes is the language of signs and construction of optimum algorithms of recognition spares so much attention?

The first reason is in that these tasks comparatively are easily added a formal and analytical decision, that and determines their attractiveness for researchers. The second reason consists in because considerable part of researchers limits the activity only theoretical reasoning's. Third problem in because the traditional is considered, that the systems of recognition are autonomous. It proves to be correct in some partial tasks, although in general case such formulation of question is not legitimate. In fact and in the systems of technical and medical diagnostics, in automated control the system by a production, recognition of defects of mechanisms and machines, determination of diagnosis of patient, recognition of the difficult hesitating modes, classification of production situations, is not an end in itself. Their recognition is needed for the receipt of initial information for the administrative system with the purpose of acceptance of leading decisions, adequate the results of recognition of unknown objects, phenomena, situations, states

It is possible to assert that reliable recognition of situations is not a sufficient condition potentially possible efficiency of control the system. But this is a necessary condition. It is difficult to present, that a

doctor which put a wrong diagnosis will find the correct method of treatment; whether the not exposure of the unsteady hesitating modes will provide reliable work of technical device.

For the procedure of measuring of primary signs must be preceded a choice or creation of criteria of recognition about the process of recognition, establishment of priority of these signs and their influence on integral descriptions of the probed process or object.

1. APPROACH TO ESTABLISHMENT OF SIGNS PRIORITY

Any classification is based on certain classification principle. In recognition theory in quality of classification principle it is possible to utilize properties of information.

Simple recognition system – homogeneous information is utilized physically (signs have the unique physical nature). For example, reading automats. In fact signs of working dictionary are the linear sizes of text. Automats are for the exchange of chinks (weight). Automats are for rejection of details (or sizes, or weight).

Difficult recognition system – heterogeneous information is utilized physically. For example, systems of medical diagnostics (analyses, cardiograms, temperature, pressure). Recognition system for geological secret service are physical and chemical properties. If in quality of principle of classification to choose the method of receipt of a posteriori information, the difficult systems are divided by an onelevel and multilevel.

The onelevel difficult systems are a posteriori information about the signs of objects which are recognized determined the direct measuring directly on the basis of treatment of results of experiments.

The multilevel difficult systems are a posteriori information about signs, determined on the basis of the indirect measuring, as a result of functioning of auxiliary recognition system.

If in quality of principle of classification to choose a priori initial information content, recognition system is divided by the systems without studies, which study with self-training. For the multilevel systems such distributing is not synonymous.

The systems are without teaching. Initial information sufficiently in order that on select principle of classification to divide all of plural of objects into classes Ω_i, $i=1,2,...,m$, to define the dictionary of signs of K_j, $j=1,2,...,N$, and on the basis of the direct working of weekend of information to describe every class of objects in language of signs.

Systems with teaching. Initial information sufficiently in order that for select principle of classification to divide all of plural of objects into classes Ω_i, $i=1,2,...,m$, to define the dictionary of signs of K_j, $j=1,2,...,N$, but not enough, to describe classes in language of signs.

The systems are with self-training. A priori initial information sufficiently for determination of dictionary of signs of $X_1, X_2, ..., X_N$. But not enough for conducting of classification, in place of pointing what class an object belongs to, the set of rules is given, in accordance with what RS makes classification. A priori information absents must be produced in the process of studies or self-training.

Regardless of on what principle the system of recognition is built it is important to set priority of the chosen signs, as authenticity of recognition depends on quality of decision of this task. At establishment of priority of signs it follows to go out from such considering:

– informative descriptions of sign are about the object of recognition;

– expenses are on determination of quantitative estimations of sign;

– influence of concrete sign is on authenticity of recognition;

– the use of method of progressive approximations is to establishment of priority.

Task of choice of the most informing subsystem of signs from some initial system – the task of objects recognition theory is important, because:

– reduction of number of signs diminishes different losses, including the losses of useful information;

– reduction of number of signs leads, as a rule, to the improvement of quality of decision [6];

– there is a limit on the number of signs which a deciding rule can be founded on at the fixed sample size. Yes, for quadratic functions and selection of $N=100$, to take rule governed guilty to contain 10 signs or more.

Let $X=\{x_1, x_2, ..., x_n\}$ it is the virgin system of signs.

It is needed to define the most informing subsystem from m signs $(m<n)$ from the point of view some criterion of F.

Will consider the subsystem of $V=\{x_{t_1}, x_{t_2}, ..., x_{t_m}\}$, $1\le t_1\le t_2\le ...\le t_m\le n$.

Will enter n-dimension vector of $U=\{u_1, u_2, ..., u_n\}$. Then subsystem V it is possible to put in accordance of $U=\{u_1, u_2, ..., u_j, ..., u_n\}$ so, that $u_j=1$, if $x_j\in V$ and $u_j=0$, if $x_j\notin V$.

A task of choice of the most informing subsystem of signs is an extreme task on a single hypercube with possible by part number of tops of C_n^m.

$F(u)$ – multiextreme analytical expression of $F(u)$ is unknown.

Examples of task of $F(u)$ are in ORT for the most widespread classes of deciding rules. For the construction of deciding rule a selection is needed (educational sequence). Let this will be the table of

$$V_p=\{X_{ij}^p\},\ p=1,2,...,k,\ j=1,2,...,n,\ i=1,2,...,N.$$

$$\sum_{p=1}^{k} V_p = N,$$

where p – a number of appearance; k – an amount of appearances; V – a sample size; N – a number of signs.

$\forall$ subsystem with m signs $\{x_{t_1}, ..., x_{t_m}\}$. With use a table V a some system is selected

$$V_p = \left\{ X_{it}^p \right\},\ t = t_1, t_2, ..., t_m,\ p = 1,2,...,k\ .$$

V_p is a total number of errors of recognition on an educational selection at the use of some fixed class of deciding rules $F(u) = \alpha(u)$.

2. ALGORITHM OF TASK DECISION

Possibly, that possibility to increase efficiency of the system of recognition due to the increase of distance between objects from different classes is. Then at implementation of some terms on middle square variation of objects into a class $S(\Omega_k)$ a task is taken to finding of maximal value for all μ from a minimum to functional $R(\Omega_k, \Omega_r)$, that it is needed to find $\max\_\min R(\Omega_k, \Omega_r)$ at limitations $\mu\ \ k, r = 1,...,n$:

$$S(\Omega_k) \le S_k^*,\quad \sum_{j=1}^{N} C_j \mu_j \le C^*\ .$$

Let $L = C_n^2 = n(n-1)/2$ is an amount of possible pair from n classes, the square of distance between the pair of classes is evened

$$R^2(\Omega_k, \Omega_r) = \sum_{j=1}^{N} \mu_j \left\{ 1/m_k\ /\ m_r \sum_{j=1}^{m_k} \sum_{l=1}^{m_r} (x_{kj}{}^i - x_{lp}{}^i)^2 \right\},$$

where $k, r = 1,...,n$.

Then, entering informing of i sign for the s pair of classes

$$\sigma_{si} = 1/m_k\ /\ m_r \sum_{j=1}^{m_k} \sum_{l=1}^{m_r} (x_{kj}{}^i - x_{lp}{}^i)^2,$$

$$s = 1,..., L,\ \ i = 1,..., N,$$

where the square of distance for a s pair is evened $R_s^2 \sum_{j=1}^{N} \mu_j \sigma_{sj},\ \ s = 1,..., L$, will get a next task

$$\max_{\mu_j = 0,1} \min_{1 \le s \le L} R_s^2 = \max_{\mu \in B} \min_{1 \le s \le L} (\mu, \sigma_s)$$

$$(C, \mu) = \sum_{1=1}^{N} \mu_j C_j < C^*,$$

where $\sigma_s = (\sigma_{s1}, \sigma_{s2}, ..., \sigma_{sl}), \mu = (\mu_1, \mu_2, ..., \mu_N)$.

In investigation of continuity of function $\min_{s=1,...,L} (\mu, \sigma_s)$, and also to simple maxmin have the method of report of maxmin, that just the following assertion:

$$\max\ \min(\mu, \sigma_s) = \lim\ \max \Psi(\mu, \sigma),$$

where $D = L_1 \times [0, \max \sum_{j=1}^{N} \sigma_{sj}]$ is cartesian product N-dimension a cube and segment $[0, \max \sum_{j=1}^{N} \sigma_{sj}]$. It is possible to take in quality a function $\underset{s=1,L}{\Psi}(\mu, v)$, for example, function

$$\Psi(\mu, v) = v - k \sum_{s=1}^{L} \left[(\mu, \sigma_s) + kG(\mu) - v - \left| (\mu, \sigma_s) + kG(\mu) - v \right| \right]^2,$$

where

$$G(\mu) = \sum_{j=1}^{N} (\mu_j{}^2 - \mu_j) - \left[C^* - (\mu, C) - \left| C^* - (\mu, C) \right| \right]^2,$$

and μ examined on the N-dimension single cube of $L_1 = \left\{ \mu : 0 \le \mu_j \le 1; j = 1,..., N \right\}$. It is undifficult to show that $\forall \varepsilon > 0 \exists K^*$ such, that $\forall k > K^*$ k just inequality

$$\max_{\mu \in B} \min_{s=1,L} (\mu, \sigma_s) - \min_{s} (\mu_{\max}(k), \sigma_s) < \varepsilon$$

Will consider a case, when except for a priori information on the location of objects in N–dimension of space and values of signs of objects the known and a priori probability of appearance of objects from the classes Ω_k, Ω_r of $k, r = 1,...,n$, that determine $P(\Omega_k)$ and $P(\Omega_r)$.

Then, assuming independence of these events, a priori probability of appearance of objects of both classes

$$P_s = P(\Omega_k) * P(\Omega_r),\ \ s = 1,..., L,\ \ and\ \ L = C_n^2\ .$$

Now a task can be formulated as follows: to find such set of working dictionary of signs, which will provide the maximal value of the expected value

$$M[R_s^2] = \sum_{s=1}^{L} R_s{}^2 * P_s$$

at limits on resources

$$(C, \mu) = \sum_{1=1}^{N} \mu_j C_j < C^*.$$

On condition that efficiency of the system of recognition can be attained only due to the increase of relation of distance between classes to middle square variation of objects into classes, the functional Φ can look like

$$\Phi = R^2(\Omega_k, \Omega_r) / S(\Omega_k) / S(\Omega_r).$$

In some cases the task of finding of working dictionary of signs succeeds to be untied due to more compact location of objects of every class, that by minimization of maximal variation of objects into a class. In this case it is possible to choose functional in a kind $\Phi = S(\Omega_k),\ \ k = 1,...,n$.

3. METHOD FOR CORRECTION PRIORITY PRIMARY SIGNS

Let's consider the algorithm for improving vocabulary of signs by the sequential removal of signs. The basis algorithm deterministic priority signs recognition is the process of establishing recognition error in sequential withdrawal symptoms.

Let's consider the process of recognition. Suppose that we have an object with the following set of features x=$\{x_1, x_2, ...x_n\}$, working dictionary that contains classes and class objects with descriptions of all the features of these classes $(\Omega_i, i = \overline{1,B}, \omega_{ij}, i = \overline{1,m}, j = \overline{1,k}, a_i, i = \overline{1,b})$.Let's take a recognition process using a distance function. This simple classification method is a very effective tool for solving such problems in which classes are characterized by degree changes limited to certain limits. Suppose that each class can be described by not the only but by multiple reference images, that is, any image that belongs to a class tends to group around one of the benchmarks $z_i^1, z_i^2, ..., z_i^{N_i}$, where N_i – number of

reference images that define the *i*-th class. Function, which determines the distance between an arbitrary image x and class ω_i will look as:

$$D = \min_l \left\| x - z_i^i \right\|, \ l = 1,2,...,N_i;$$

It means that D_i is the smallest of the distances from each image to class ω_i standard. Let's compute the value of distances D_i, $i = 1,2,...,M$, and classifying image belongs to the class ω_i if the condition is $D_i < D_j$ fair for all $j \neq i$. In case of equal distances decision is made randomly.

In case of equal distances decision is made randomly.

The next step is the establishment of the recognition error. Ideally, the error = 0, the characteristic value of the investigated object would match the average values with feature vector class:

$$k_s = \frac{x_{i,2} - x_{i,1}}{2}.$$

But it is quite rare, so the conditional interpretation that the average grade point is ideal and error estimation at these points = 0 is introduced. Instead deviation values of attributes of the object will be closer to 50%, this is due to the fact that the working dictionary attributes the difference between the features of different classes can be "1". Therefore:

$$S_s = \sum_i (k_{i,2} - k_s)^2;$$

Error estimating can be presented:

$$K = \frac{S_s}{50}.$$

The error recognition process is equal to:

$$P = 50 - \frac{S_{\min}}{K},$$

where $S_{\min}$ – the minimum value of properties belonging to a known class.

The next step is the removing if one of the signs performing the recognition process with maintaining all the results. This process is carried out until all the symptoms of the object are disguised.

The next step is the comparison of the results of alternate recognition without one feature and pre result taking into account all attributes. First the signs, the absence of which changed the outcome of the test object class membership ω_i. are selected. They will be more informative. The next step will be the selection of the signs that led to controversial situation. They will have the medium priority, because their absence causes a partial error (50%) of belonging to a certain class. Signs that remain will be understood as little information as their absence does not affect the result of recognition (the error is 0%).

Identify of the priorities is adding new features to the selected base. This method is rather cumbersome, because it would include $\frac{(n-1)!}{2}$ operations, while only preliminary (n+1) operations. To reduce the number of operations we should consider combining those features which give absence of 100 or 50% error. It will reduce the number of operations $\frac{(t-1)!}{2}$ where $t << n$ much. Algorithm of realization is similar to sorting algorithm in sequential shutdown of one trait.

Thus, we can identify the most informative features and clarify the working vocabulary of signs. The significant solving of this problem is extremely important, as it is affecting on the efficiency of the recognition system and is appropriate in decision management.

Based on the proposed algorithm, we can select the priority signs as the most informative features and clarify the working vocabulary of signs. The significant solving of this problem is extremely important, as it is affecting on the efficiency of the recognition system and is appropriate in decision management.

4. RESULTS OF DESIGN COMPUTER

That it is expedient to estimate the value of problem of recognition, it is enough to say, that creation of artificial intelligence not as imitations of intellection of man, and a recreation of thinking process is creation of the system of recognition, which after the possibilities approaches possibilities of human brain.

Efficiency of recognition sure depends on the rightness of choice of working dictionary of signs. Than signs are better chosen the probed object is determined more precisely. Usually signs will have the appearance vector (accordingly possible max and min value). Desired in signs are noticeable enough differences between classes, as near enough values can result in the failures of the system and wrong taking of object to the certain class. In such case more effective will be other signs which have more expressed differences, they will have greater priority. It is also important to mark that the size of vector of signs has also a ponderable influence on priority of signs. Accordingly than greater vector of signs, the less priority of sign and vice versa than less vector of sign, the its greater priority. In that case when the size of vector is large (for example, a bicycle has from 4 to a 1 wheel) – the value of the probed object can hesitate in large limits (from max to min). But when a size of vector of sign is small (for example, a machine has 4 wheels), in such case there is exact determination of object and absence of which worsens the process of recognition.

CONCLUSIONS

1. The working dictionary of signs, which are utilized both for a priori description of classes of objects and for a posteriori description of unknown objects which act on the entrance of the system of recognition, it is needed to take into account that even if in relation to some signs a priori information, sufficient for description of classes of objects in language of these signs, can be got, in a working dictionary it is needed to include the most informing signs only.

2. The task of choice of optimum working dictionary of signs is taken to the task of the linear or nonlinear programming. The choice of that or other method of decision of task strongly depends from properties of functional of quality, which can get out on different.

3. At the decision of task in the scholastic raising, that when a priori probabilities of appearance of objects of classes, value are known criteria of efficiency of the system of recognition considerably higher than in determined case.

4. The estimations of expedience of the use are got each of signs and its influence on authenticity of recognition.

5. The results of computer design confirmed legitimacy of the use of the resulted estimations to the construction of the systems recognitions which are based on the use of the determined signs.

6. The resulted results allow optimizing procedure of constructing of algorithms of recognition of objects with a difficult dynamics, and also objects which are described on the basis of the determined and probabilistic signs.

7. After several experiments it was found that those signs which will include less of values will be more informing.

REFERENCES

[1] *Горелик А.Л., Скрипник В.А.* Методы распознавания. — М.: Высшая школа, 1989. — 232 с.

[2] *Заяць В.М.* Дискретні моделі коливних систем для аналізу їх динаміки. Монографія. — Львів; В-во УАД. — 2011. — 284 с.

[3] *Заяць В.М.* Комп'ютерна система ідентифікації особи за її рукомоторними реакціями // Праці Міжнар. конф. „Штучний інтелект". – Том 2. — Крим. – 2002. — С. 214-218.

[4] *Заяць В.М., Камінський Р.М.* Методи розпізнавання образів. Навчальний посібник. – Львів: „Львівська політехніка". – 2004. — 176 с.

[5] *Заяць В.М., Заяць М.М.* Математичний опис системи розпізнавання користувача комп'ютера // Зб. "Фізико-математичне моделювання та інформаційні технології". – Львів. – 2005. – Вип. 1. — С. 146-152.

[6] *Заяць В.М., Іванов Д.О.* Архітектура подіє-орієнтованих систем на прикладі системи розпізнавання рукописного тексту // Вісник НУ „Львівська політехніка" „Комп'ютерна інженерія та інформаційні технології". – Львів. – 2004. — № 530. — С. 78-83.

[7] *Фукунага К.* Введение в статистическую теорию распознавания. — М.: Наука, 1979. — 512 с.

[8] Харкевич А.А. Опознание образов // Радиотехника. — 1959. — Том 14. — С. 15-19.

[9] *Шокира Г.Я.* Застосування методу послідовних наближень для встановлення пріоритету первинних ознак / Г.Я. Шокира // Математичне моделювання складних систем. – Львів, ЛДІНТУ. — 2009. — С. 9-10.

[10] *Zayats V., Ivanov D.* Structural method of hand-written text recognition // Pros. International Conf. "The experience of designing and application of CAD systems in microelectronics". – Lviv-Polyana. — 2005. – P. 493-494.

Manuscript received June, 4, 2012

Vasyl Zayats, DSc, professor of Lviv National Polytechnic University. Research interests: computer and mathematical modeling and recognition and identification systems difficult nature.

Galina Shokyra, post graduate student of Lviv National Polytechnic University. Research interests: recognition and identification systems with determine and probalistic signs.

УДК 519:616-079.4: 621. 372:621.314

Определение приоритета первичных признаков и его коррекция при построении систем распознавания обьектов / В.М Заяць, Г.Я. Шокира // Прикладная радиоэлектроника: науч.-техн. журнал. – 2012. – Том 11. № 3. – С. 361–365.

Предложен подход к установлению приоритета первичных признаков при построении систем распознавания на основе метода последовательных приближений. Разработан метод коррекции установленных приоритетов путем добавления новых признаков к выбранному неполному базису первичных признаков. Полученные результаты подтверждены постановкой компьютерных экспериментов при моделирования элементов систем распознавания с детерминированными признаками

Ключевые слова: система распознания, первичные признаки, рабочий словарь, приоритет, метод коррекции.

Библиогр.: 10 назв.

УДК 519:616-079.4: 621. 372:621.314

Визначення пріоритету первинних ознак і його корекція при побудові систем розпізнавання об'єктів / В.М. Заяць, Г.Я. Шокира // Прикладна радіоелектроніка: наук.-техн. журнал. – 2012. – Том 11. № 3. – С. 361–365.

Запропоновано підхід до встановлення пріоритету первинних ознак при побудові систем розпізнавання на основі методу послідовних наближень. Розроблено метод корекції встановлених пріоритетів шляхом додавання нових ознак до вибраного неповного базису первинних ознак. Отримані результати підтверджені постановкою комп'ютерних експериментів при моделювання елементів систем розпізнавання з детермінованими ознаками.

Ключові слова: система розпізнання, первинні ознаки, робочий словник, пріоритет, метод корекції.

Бібліогр.: 10 найм.