UDK 004.827

D. E. DVUKHGLAVOV, O. V. MUZYKA, S. O. HLAZKOV

MODEL OF THE SITUATIONS RECOGNITION IN CONDITIONS DISSIMILAR AND INCOMPLETE DATA

There have been proposed model of situations recognition in determined alphabet, based on combination of quantitative and qualitative characteristics, considered the data dissimilar, coming from information sources. In the model improved the quality of the formalized description of quantitative characteristics using histograms instead of fuzzy L-R intervals, which gives the possibility of organizing a self-learning recognition system situations based on the processing statistics of recognition. Recommendations for formation of characteristics sets to overcome the data incompleteness have been present. Proposed model may be use for formalization knowledge about situation recognition process. Subject areas for implementation taken results are diagnostic in medicine and energetic, situation assessment at the military control points, making prognosis in economic.

Keywords: situations recognition, alphabet of classes, quantitative characteristic, qualitative characteristic, fuzzy measure, incomplete data

Представляється модель розпізнавання ситуацій у детермінованому алфавіті на основі сукупності кількісних та якісних ознак, що враховує різнорідність даних, що поступають від джерел інформації. В моделі вдосконалюється якість формалізованого опису кількісних ознак за рахунок використання гістограм заміст нечітких L-R інтервалів, що надає можливість організації самонавчання системи розпізнавання ситуацій на основі обробки статистики розпізнавання. Представлені рекомендації щодо формування наборів ознак для подолання неповноти даних.

Ключові слова: розпізнавання ситуацій, алфавіт класів, кількісна ознака, якісна ознака, нечітка міра, неповнота даних.

Представляется модель распознавания ситуаций в детерминированном алфавите на основе совокупности количественных и качественных признаков, учитывает разнородность данных, поступающих от источников информации. В модели усовершенствуется качество формализованного описания количественных признаков за счет использования гистограмм вместо нечетких L-R интервалов, что позволяет организовать самообучение системы распознавания ситуаций на основе обработки статистики распознавания. Представлены рекомендации по формированию наборов признаков для преодоления неполноты данных.

Ключевые слова: распознавание ситуаций, алфавит классов, количественный признак, качественный признак, нечеткая мера, неполнота данных.

Introduction. At the present stage of management processes automation principles development more often automation system execute analytical information processing, which necessary for effective decisions making. In set of these task also must be include recognition situations task.

The theoretical basis of recognition situations present in [1]. Modern recognition situations methods base on fuzzy sets [2], functionality ordinal networks [3], Bayesian approach [4] and neural networks [5]. The difficulty of using some detection methods, in particular Bayesian approach and neural networks is that their application requires the collection of statistics over a long period. To obtain such a set is not always possible. A portion of the methods works only with the characteristics of situations, which are continuous variables. Therefore, more and more to recognize situations relies on methods that based on the formalization of expert knowledge. At the same time must retain the possibility of training the system to improve recognition quality. It can argued that the methods of pattern recognition, which fully satisfy modern requirements for this class of systems does not exist.

In our time actuality of decision this tasks we can see in medicine [6], in military affairs [3, 7], also in robotics, in automatic control, during economic analysis. Research in this scientific area, along with research in the field of pattern recognition, correspond introduction of intellectual data processing methods in control points activity, which is an additional evidence of this research area actuality.

The purpose of the article is development of method, which allows to represent data for identification of different nature and to ensure situations recognition in conditions of incomplete data.

Resolution of the situations recognition task. The situations (or objects) recognition task is to make conclusion about the class of observed situation (objects) by analyzing their characteristics (parameters and relationships with other objects subject area).

The concept of "class" meets a set of observed situations (objects), characterized by the regularities of manifestation properties.

The features of the observed situation, presented in recognition system as set of characteristics $\{X_1,...,X_L\}$, where L – number of characteristics used to identify situations of a certain type. The list of features is common to all type situations, but these symptoms manifest themselves differently in different classes of situations.

The observed situations classes grouped in the alphabet. The alphabet is disjoint set of classes:

$$A_m = \{K_1^m, K_2^m, ..., K_{M_m}^m\}, \quad (1)$$

where K_i^m – classes of alphabet A_m;

M_m – number of alphabet classes.

Typical tasks of this type is to determine the general condition based on initial examination of patients based survey in foster recognition department or type of airplane or ship according to the various technical means.

The formalization of disparate source data. Based on the analysis of characteristics used in the detection, we can conclude that their composition is diverse in terms of mathematical properties values.

In many cases, the set of possible values of the characteristic is a subset of the real numbers:

$$X_k \in [X_{k_{\min}}, X_{k_{\max}}], \quad (2)$$

where X_k – *k*-th characteristic used to identify classes of objects in the *m*-th alphabet;

$X_{k_{\min}}$, $X_{k_{\max}}$ – minimum and maximum possible values of attributes for objects that are recognized.

Such characteristics called *quantitative*. For example, for the recognition of aircraft objects types such characteristics are speed and altitude, in medical diagnostic – are temperature and human height.

Characteristics that take value from a specified list and do not have the structure in any order, called *quality*:

$$X_k \in \{z_1^k, z_2^k, ..., z_{Z_k}^k\}, \quad (3)$$

where z_p^k – *p*-th possible value of *k*-th characteristics used to identify classes of objects in the *m*-th alphabet;

Z_k – total number of possible values of this characteristic.

The assignment to a class of situations involves determining measures of proximity of the object observed for each of the classes $val(K_i^m)$ and the decision to situations class selected in accordance with the rules.

For this type of characteristics in solving the problem of recognition of air objects include the size of the group aircraft or the nature of the maneuver; in medicine – the overall functional state or the nature of employment.

The basis for constructing rules of measures of proximity of objects observed classes alphabet is to define the set of alternative groups characteristics, patterns of manifestation which will be analyzed. Formally, this can be determined by the following expression:

$$K_i^m = \bigcup_{q=1}^{Q_i} G_q = \bigcup_{q=1}^{Q_i} \bigcap_{k \in R_{iq}} X_k, \quad (4)$$

where G_q – alternative group characteristics q, which allow to make recognization;

Q_i – number of alternative groups characteristics used to identify *i*-th class situation;

R_{iq} – set of indexes of *q*-th alternative characteristics group, used to identify the *i*-th class situations.

For each characteristic, which used for recognition *i*-th class situations as part of *q*-th group, experts determined the pattern of values characteristics appearance for a particular situations as fuzzy set N_k^{miq}, which can be presented as

$$N_k^{miq} = \{(X_k, \mu^{miq}(X_k))\}, \quad (5)$$

where $\mu^{miq}(X_k)$ – degree of belonging to *i*-th class situations according to values of characteristic X_k^m as part of *q*-th alternative group of characteristic.

One approach to knowledge representation pattern of display values in the form of quantitative characters is the use of histograms. This presentation simple to understand experts allows a satisfactory degree of reliability to pass basic laws of the subject area and is easy in processing.

To construct a histogram certain features expert must set the width of the interval R^k or their number K^k. In the last case the width of the interval is calculated according to expression:

$$R^k = \frac{X_{k_{\max}} - X_{k_{\min}}}{K^k}. \quad (6)$$

For each interval of histogram expert must determine the value of belonging situation to a certain class. Membership function values for of quantitative characters takes shape:

$$\mu^{miq}(X_k) = \begin{cases} \mu_1^{miqk} \middle| X_k \in \left[X_{k_{\min}}, X_{k_{\min}} + R_{inm}^k\right[; \\ \mu_2^{miqk} \middle| X_k \in \left[X_{k_{\min}} + R_{inm}^k, X_{k_{\min}} + 2 * R_{inm}^k\right[; \\ \dots \\ \mu_{K_{inm}^k}^{miqk} \middle| X_k \in \left[X_{k_{maa}} - R_{inm}^k, X_{k_{maa}}\right], \end{cases} \quad (7)$$

where μ_s^{miqk} – the function belonging situation to the *i*-th class of alphabet *m*, if the value of characteristic X_k consisting of *q*-th group is the interval *s*.

Graphic illustration representing regularities of quantitative characters values shown in fig. 1.

In [3] presents another well-known version of the presentation patterns of display characteristics is the use of fussy L-R intervals – trapezoidal functions with key points (A, 0), (B, 1), (C, 1), (D, 0), where A and D correspond to boundaries of the possible values of the trait, B, and C– the range border of most possible values.

It exist possibility this model representation of quantitative characteristics in the model based on the histograms. This requires the respective intervals of the histogram to assign the degree of conformity the same intervals of the trapezoid used in model of L-R intervals. However, the use of the proposed model for formalization of quantitative traits allows to consider more complex patterns of onset, and, therefore, to provide more accurate situation recognition.

To formalize of the appearance regularities of quality attributes in the objects of a certain class expert must determine the extent possible assignment object to a class for all possible values of certain properties. Membership function values on quality characteristics shall look as present below (expression (9)):

$$\mu^{miq}(X_k) = \begin{cases} \mu_1^{miqk} \middle| X_k^m = z_1^{mk}; \\ \mu_2^{miqk} \middle| X_k^m = z_2^{mk}; \\ \dots \\ \mu_{Z_k}^{miqk} \middle| X_k = z_{Z_k}^k. \end{cases} \quad (9)$$

Graphic illustration representing regularities of quality characteristics values presented in fig. 2.

Common list of alphabet classes list of characters on which the identification and description of regularities of characters in object form a formalized description of the authentication process.

A description of the appearance regularities of characteristics in classes of a given alphabet is a formalization of the initial stage of recognition. For further situations recognition it is necessary to establish rules for attributing to a particular class of situations, depending on the specific values of characteristics. In [3] a description of these rules is proposed to be based on AND-OR graph. In fact, this graph corresponds to a Boolean expression. But any logical expression, if it is not a logical contradiction, can be represented in disjunctive normal form. This approach proposes to use, i.e. features will be divided into groups and the class situation will be identified at identification of compliance by all indications the group.

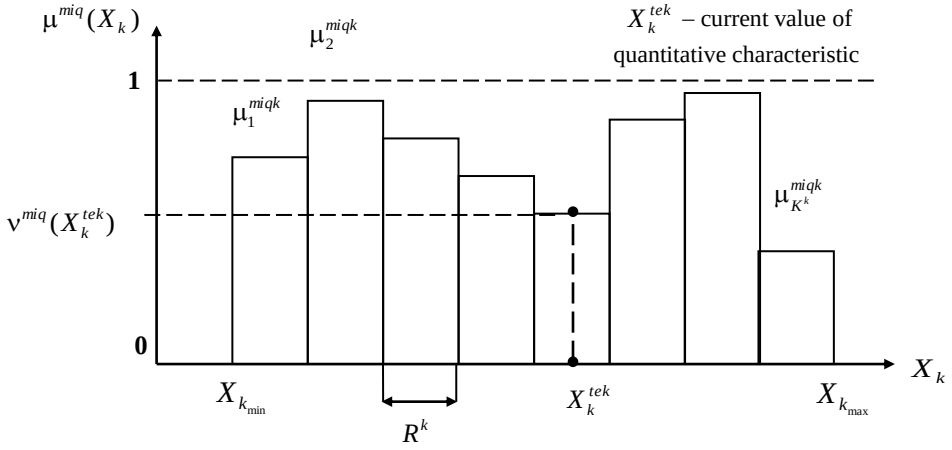

Fig. 1 – Graphic illustration of the formalization of the laws of manifestation quantitative characteristics values using histograms

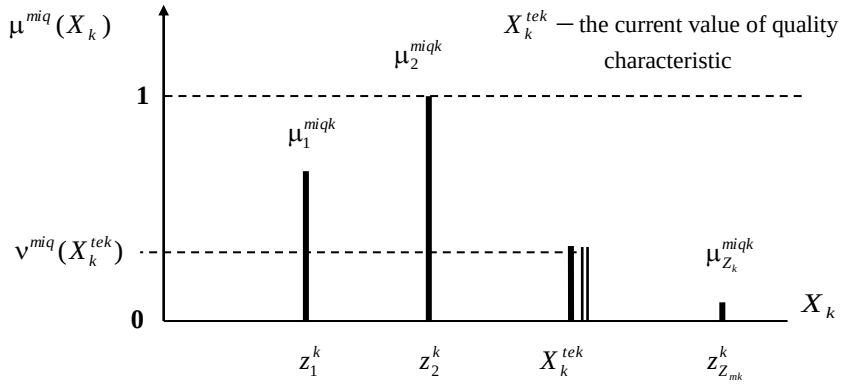

Fig. 2 – Graphic illustration formalize patterns of display values of quality features

The recognition procedure. The class definition of the object observed made in the following order:

getting a set $\{X_1^{tek},\ldots,X_L^{tek}\}$ – set of characteristics values of particular observed situation;

determining extent of truth $\nu^{miq}(X_k^{tek})$, which show accordance the current value of characteristic X_k to experts describe patterns of their appearance in situations in the *i*-th class of *m*-th alphabet, when characteristic looking as part of *q*-th alternative group of characteristic;

determining measures $\omega^{mi}(G_q)$, describing the proximity of the situation at to the *i*-th class of *m*-th alphabet on the results determine the extent of truth of alternative signs G_q;

determining generalized measures of proximity situation $\omega(X_k^m)$ to classes of *m*-th alphabet;

determining class K^{m*} of situation according to the set rules of decision making about the class object.

The value of the signs X_k^{tek} information have been taken from sources in the form of estimates, which are for quantitative traits – point estimation of their values, for signs of quality –possible values of a given set.

Evaluation of truth quantitative trait determined according to expression:

$$\nu^{miq}(X_k^{tek}) = \mu_s^{miqk} \left| X_k^{tek} \in \left[\begin{matrix} X_{k_{\min}}^m + (s-1) * R_{inm}^k \\ X_{k_{\min}}^m + s * R_{inm}^k \end{matrix} \right[\right. . \quad (10)$$

ISSN 2079-0023 (print)
ISSN 2410-2857 (online)

Evaluation of truth qualitative characteristics will be determined in accordance with rule:

$$\nu^{miq}(X_k^{tek}) = \mu_s^{miqk} \Big| X_k^{tek} = z_s^k. \qquad (11)$$

For further recognition, the following equality, just as the theory of fuzzy sets:

$$\nu(X_1 \cap X_2) = \min\left\{\nu(X_1), \nu(X_2)\right\}; \qquad (12)$$

$$\nu(X_1 \cup X_2) = \max\left\{\nu(X_1), \nu(X_2)\right\}. \qquad (13)$$

Defining measures proximity object to the *i*-th grade m-th alphabet degrees of truth for signs of alternative G_q carried out according to the rules of crossing the fuzzy sets (expression (12)). The result is:

$$\omega^{mi}(G_q) = \min_{k \in R_{iq}} \{\nu^{miq}(X_k^{tek})\}. \qquad (14)$$

Defining measures proximity object classes alphabet is made by applying the rules of association of fuzzy sets (expression (13)). The merger assessments proximity of the object to a class, the value is:

$$\omega(K_i^m) = \max_q \{\omega^{mi}(G_q)\}. \qquad (15)$$

The basic rule is used to make decisions about the class object of observation is the following:

$$K^{m^*} = \arg\max_i \{\omega(K_i^m)\}. \qquad (16)$$

Under this rule class of situation is one measure which is close to the maximum.

To ensure the required probability recognition results using a rule the following form:

$$K^{m^*} = \arg\max_i \{\omega(K_i^m)\} \;\Big|\; \omega(K_i^m) \geq \delta. \qquad (17)$$

According to this rule in determining the class of sutiation are considered, a measure of intimacy which exaggerates set to δ. The value threshold δ lies in the range [0, 1].

While the decision to object class may be a situation where a measure close to one object slightly exaggerates the extent of that proximity of the object to another class. If necessary, prohibition decisions on class facility in such conditions can be used the following rule:

$$\exists i \forall j \left[\begin{array}{c} ((\omega(K_i^m) - \omega(K_i^m) > \Delta) \wedge (i \neq j)) \rightarrow \\ (K^{m^*} = K_i^m) \end{array} \right]. \qquad (18)$$

The value Δ called typical margin. Its importance also are in the range [0, 1]. Also must note, when $\Delta = 0$, will be taken rule, represented by expression (17).

In practice can be applied all at once strategy.

In case of dissatisfaction with the requirements of the rules of decision-making about the class object carried refusal of recognition.

Conclusions. Results of the research question of formalizing recognition process suggest as follows:

dissimilar of data sources appears in traits nature – they can be evaluated either quantitative or qualitative; the development of the recognition system must be treated mentioned traits in both groups;

incomplete data can be overcome through the analysis of the situation or object from some positions, which increases the amount of formalized description, but increases the likelihood of obtaining results;

use histograms to represent quantitative traits is a good alternative to the use of L-R intervals, providing the possibility of self-recognition system as the accumulation of relevant statistics.

Presents the results and findings are the basic for development, testing and introduction into practice of corresponding personal recognition situation subsystem software. Important for the implementation of the model is to develop and implement self-learning model algorithm . In turn, this task makes it relevant to the question of comparative analysis of work procedures for the recognition of situations on the basis of the proposed model with the procedures of recognition based on Bayesian approach and based on neural networks.

References

1. *Горелик А. Л.* Методы распознавания / *А. Л. Горелик, В. А. Скрипкин.* – М. : Высшая школа, 1984. – 207 с.
2. *Вятченин Д. А.* Нечеткие методы автоматической классификации / *Д. А. Вятченин*. – Минск : УП «Технопринт», 2004 – 219 с.
3. *Грачев В. М.* Методика распознавания классов воздушных объектов в АСУ ПВО с использованием однородной функциональной сети / *В. М. Грачев, А. Н. Попрыгин* // Сб. научн. тр. ХВУ. – Х. : ХВУ. – 1995. – Вып. 8. – С. 49–54.
4. *Вагис А.* Эффективность байесовских процедур распознавания / *А. Вагис., А. Гупал* // ITHEA International Scientific Socaety – 2008. – Available at : http://www.foibg.com/ibs_isc/ibs-15/ibs-15-p11.pdf. – Accessed : 20th of October 2016.
5. *Барский А. Б.* Нейронные сети: распознавание, управление, принятие решений / *А. Б. Барский*. – М. : Финансы и статистика, 2004. – 176 с.
6. *Голоскоков А. Є.* Процедура діагностування стану сердцево-судинної системи пацієнту на основі нечіткої логіки / *А. Є. Голоскоков, К. В. Мельнік* // Вісник НТУ «ХПІ». Тематичний випуск: Інформатика і моделювання. – Х. : НТУ «ХПІ». – 2008. – № 49. – С. 101–104.
7. *Павленко М. А.* Метод формализации знаний о процессе распознавания ситуацій нарушения правил движения воздушными судами / *М.А. Павленко* // Системи управління, навігації і звя'зку. – К. : ДП «ЦНДІ НіУ», 2012. – Вип. 2 (22). – С. 86–92.

References (transliterated)

1. Gorelik A. L., Skripkin V. A. *Metody raspoznavanija* [The Recognization Methods]. Moscow, Vysshaja Shkola Publ, 1984, 207 p.
2. Vjatchenin D. A. *Nechetkie metody avtomaticheskoj klassifikacii* [Indistinct Methods of Automatic Classification]. Minsk, UP «Technoprint», 2004, 219 p.
3. Grachev V. M., Poprygin A. N. *Metodika raspoznavanija klassov vozdushnyh ob#ektov v ASU PVO s ispol'zovaniem odnorodnoj funkcional'noj seti* [Technique of air objects classes recognition in AirDefense ACS with use of uniform functional network]. *Sb. nauchn. tr. KhVU* [Collection of scientific papers of Kharkov Military University]. Kharkov, KhMU, 1995, no. 8, pp. 49–54.
4. Vagis A., Gupal A. *Jeffektivnost' bajesovskih procedur raspoznavanija* [Efficiency of Bayesian procedures of recognition]. ITHEA International Scientific Socaety, 2008. Available at: http://www.foibg.com/ibs_isc/ibs-15/ibs-15-p11.pdf (accessed 20.10.2016).

5. Barskij A. B. *Nejronnyeseti: raspoznavanie, upravlenie, prinjatiereshenij* [Neural networks: recognition, management, decision-making]. Moscow, Finansyand Statistika Publ., 2004, 176 p.
6. Goloskokov A. Ye., Mel`nik K. V. *Procedura diagnostuvannya stanu serdcevo-sudynnoyi systemy paciyentu na osnovi nechitkoyi logiky* [The procedure of diagnosing the state of patient cardiovascular system based on fuzzy logic]. *Visnyk NTU "KhPI"* [Bulletin of the National Technical University "KhPI"]. *Tematychnyy vypusk: Informatyka i modelyuvannya* [Special issue: Informatics and modeling]. Kharkov, NTU "KhPI" Publ., 2008, no. 49, pp. 101–104.
7. Pavlenko M. A. *Metod formalizacii znanij o processe raspoznavanija situacij narushenija pravil dvizhenija vozdushnymi sudami* [Method of knowledge formalization of process of recognition of the movement rules violation by aircrafts situations]. *Systemy upravlinnya, navigaciyi i zvyazku* [Control systems, navigation and communication]. Kyiv, GE «CSRI N&C», 2012, no. 2 (22), pp. 86–92.

Received 14.10.2016

Бібліографічні описи / Библиографические описания / Bibliographic descriptions

Модель розпізнавання ситуацій в умовах різнорідності та неповноти даних / Д. Е. Двухглавов, О. В. Музика, С. О. Глазков // Вісник НТУ «ХПІ». Серія: Системний аналіз, управління та інформаційні технології. – Х. : НТУ «ХПІ», 2016. – № 37 (1209). – С. 17–21. – Бібліогр. : 7 назв. – ISSN 2079-0023.

Модель распознавания ситуаций в условиях разнородности и неполноты даних / Д. Э. Двухглавов, А. В. Музыка, С. А. Глазков // Вестник НТУ «ХПИ». Серия: Системный анализ, управление и информационные технологии. – Харьков : НТУ «ХПИ», 2016. – № 37 (1209). – С. 17–21. – Библиогр. : 7 наим. – ISSN 2079-0023.

Model of the situations recognition in conditions dissimilar and incomplete data /D. E. Dvukhglavov, O. V. Muzyka, S. O. Hlazkov // Bulletinof NTU "KhPI". Series: System analysis, control and information technology. – Kharkov : NTU "KhPI", 2016. – No. 37 (1209). – P. 17–21. – Bibliogr. : 7. – ISSN 2079-0023.

Відомості про авторів / Сведения об авторах / About the Authors

Двухглавов Дмитро Едуардович – кандидат технічних наук, доцент, Національний технічний університет «Харківський політехнічний інститут», доцент кафедри «Програмна інженерія та інформаційні технології управління»; тел.: (095) 120-30-66; e-mail: ddimae72@gmail.com.

Двухглавов Дмитрий Эдуардович – кандидат технических наук, доцент, Национальный технический университет «Харьковский политехнический институт», доцент кафедры «Программная инженерия и информационные технологии управления»; тел.: (095) 120-30-66; e-mail: ddimae72@gmail.com.

Dvukhglavov Dmytro Eduardovych – Candidate of Technical Sciences (Ph. D.), Docent, National Technical University "Kharkiv Polytechnic Institute", Associate Professor at the Department «Software engineering and management information technology»; tel.: (067) 839-12-41; e-mail: ddimae72@gmail.com.

Музика Олександр Володимирович – Національний технічний університет «Харківський політехнічний інститут», студент; тел.: (067) 839-12-41; e-mail: adept1994@gmail.com.

Музыка Александр Владимирович – Национальный технический университет «Харьковский политехнический институт», студент; тел.: (067) 839-12-41; e-mail: adept1994@gmail.com.

Muzyka Oleksandr Volodymyrovych – National Technical University "Kharkiv Polytechnic Institute", student; tel.: (067) 839-12-41; e-mail: adept1994@gmail.com.

Глазков Станіслав Олександрович – Національний технічний університет «Харківський політехнічний інститут», студент; тел.: (093) 251-99-52; e-mail: ststasker@gmail.com.

Глазков Станислав Александрович – Национальный технический университет «Харьковский политехнический институт», студент; тел.: (093) 251-99-52; e-mail: ststasker@gmail.com.

Hlazkov Stanislav Oleksandrovych – National Technical University "Kharkiv Polytechnic Institute", student; tel.: (093) 251-99-52; e-mail: ststasker@gmail.com.