

A MODIFIED CASE-BASED REASONING METHOD BASED ON THE ROUGH SET THEORY

Kovalenko I. I. – Dr. Sc., Professor, Professor of the Department of Software Engineering, Petro Mohyla Black Sea National University, Mykolayiv, Ukraine.

Shved A. V. – PhD, Associate Professor of the Department of Software Engineering, Petro Mohyla Black Sea National University, Mykolayiv, Ukraine.

Koval N. V. – Post-graduate student of the Department of Software Engineering, Petro Mohyla Black Sea National University, Mykolayiv, Ukraine.

ABSTRACT

Context. Knowledge bases are the main element of artificial intelligence systems. They are formed on the basis of two generally accepted approaches: the object-oriented approach and object-structural approach. Knowledge structuring through its ordering, classification and typing of selected classes is the main operation that is implemented in both approaches. Quite often there are situations when data or knowledge is not exact and it is impossible to perform their exact classification. These features necessitate the development of new approaches aimed at solving problems of extracting knowledge from large arrays of unordered data, structuring, presenting and analytical processing of inexact knowledge in automated construction of knowledge bases.

Objective. The objective of this paper is a research of new approaches for solving problems of representation of knowledge about cases in intellectual decision support systems.

Method. An approach aimed at modifying Case-Based Reasoning method on the basis of Rough Set Approach has been proposed in this paper. The proposed method forms a partition of cases to determine the degree of their belonging to the goal classes using upper and lower approximations of goal classes, considering the relative importance of classification attributes and formed equivalence classes.

Results. The proposed modification of Case-Based Reasoning method allows extracting knowledge about cases from arrays of unordered data with the purpose of the case base construction, and handling the inconsistent (in cases where for the same values of attributes cases belong to different classes), and incomplete (in cases where the values of some attributes or information of the case belonging to the given class is missing or unreliable) information about cases.

Conclusions. The proposed method of representation knowledge about cases, their adaptation and subsequent search in the case base formed under uncertainty and existence of inexact, rough, inconsistent initial data constitutes a theoretical basis for constructing intellectual decision support systems.

KEYWORDS: Rough Set Theory, Case-Based Reasoning, case (precedent), knowledge base, case base, classification.

ABBREVIATIONS

CB is a case base;

CBR is a case-based reasoning;

DM is a decision-maker;

RST is a rough set theory;

x_i is a parameter (characteristic) of the case;

X_i is a category in universe U ;

$[x]_R$ is represented an equivalence class in R containing an element $x \in U$;

U is a finite set of considered objects (universe).

NOMENCLATURE

α_R is a degree of completeness of existing knowledge;
 ρ_R is a degree of incompleteness of existing knowledge;

A is a set of case attributes;

a_i is a value of case attributes;

$BN_R(X)$ is a R -boundary region of X ;

$card(Y)$ is a cardinality of Y ;

CX_i is a region of acceptability of the corresponding case parameters x_i ;

D are recommendations to the decision-maker;

E_j is a class of equivalence relations over U ;

$IND(R)$ is a family of all equivalence classes of R ;

K is a relational system (knowledge base);

n is a number of case parameters;

$NEG_R(X)$ is a R -negative region of X ;

$POS_R(X)$ is a R -positive region of X ;

R is a family of equivalence relations over U ;

$\underline{R}X$ is a \underline{R} -lower approximation of rough set X ;

$\overline{R}X$ is a \overline{R} -upper approximation of rough set X ;

INTRODUCTION

CBR method based on use of previous experience in decision-making became widespread recently for various problems solving. A case (precedent) is a structured representation of accumulated experience in the form of data and knowledge, ensuring its subsequent automated processing with the help of specialized software systems [1, 2].

In general, the case model is represented as follows:

$$Case(x_1, x_2, \dots, x_n, D). \quad (1)$$

In the process of CBR implementation the following basic tasks should be solved: choice of the form of knowledge representation about the case, identification and extraction of the case from the generated knowledge base (case base) and adaptation of computed solutions [3].

To solve the first problem, traditional methods of knowledge extraction and presentation are widely used,

such as commutative and textual methods, product models, semantic networks, frames, formal logic models, etc. The search and extraction of cases is performed using fuzzy sets and fuzzy relationships, decision trees, neural networks, and others.

Along with that to solve this problem method of search for the nearest neighbor and its modification became widespread. As for the task of adapting decisions based on cases, very few works have been devoted to its analysis.

The object of study is the process of structuring of case knowledge derived from large arrays of raw, rough data, by ordering and classifying them in the knowledge base using equivalence relations.

The subject of study are the models and methods of modeling reasoning on the basis of precedents, and their modifications based on the mathematical apparatus of the rough set theory.

The purpose of the work is to consider possibilities of RST for solving problems of representation knowledge about cases, their search and adaptation for subsequent decision-making.

1. PROBLEM STATEMENT

Let the initial set of knowledge about cases X be represented by two classes of cases: $Case_{S1} = \{x_1, x_2, x_3, x_4\}$ and $Case_{S2} = \{x_5, x_6, x_7\}$, where x_i is some parameter (characteristic) of the case $x_i \in CX_i$, and CX_i is a region of acceptability of the corresponding case parameters. And let $X_0 = \{x_3, x_4, x_7, x_8\}$ be some set of knowledge, which must be attributed to one of the indicated classes $Case_{S1}$ and $Case_{S2}$. However, it can be seen that elements of this set $\{x_3, x_4\} \in Case_{S1}$ and, $\{x_7\} \in Case_{S2}$, but the element $\{x_8\} \notin (Case_{S1}, Case_{S2})$. This characterizes the situation when it is not possible to perform an exact classification.

Suppose given a set of cases $SC = \{Case_{Sj}\}$, $j = 1, 2, \dots, k$, where $Case_{Sj} = \{x_i\}$, $i = 1, 2, \dots, n$, and some set of knowledge $X_0 = \{x_i\}$, $i = 1, 2, \dots, p$ ($p \geq n$).

The problem of classification can be presented as the problem of choosing the correct class $Case_{Sj}$ for an each given input element $x_i \in X_0$ from the original set of knowledge X_0 .

2 REVIEW OF THE LITERATURE

There are a lot of works devoted to models and methods of analyzing cases [1–7, etc.]. For example, in [1, 2] issues of CBR modeling in intelligent decision support systems are considered. The paper [3] is devoted to the description of case extraction approaches using fuzzy sets and fuzzy relations. In [4, 5] issues of practical application of the nearest neighbor search method and its modifications are considered as well as problems connected with its use in the task of cases extracting are noted. The analysis of listed publications allows for the conclusion that existing methods of representation and structuring of knowledge about cases use strict models determined by the relevance (certainty) of considered elements.

At the same time in real situations it is quite often necessary to solve problems of extracting knowledge from arrays of unordered (raw, rough) data. The knowledge thus obtained is not accurate and it is impossible to perform its exact classification (establish a classification category). This is mainly due to the fact that the “rigidity” of existing models of knowledge representation forces developers to combine or cut down the real knowledge of experts [3].

To analyze such situations the author of [8] has proposed RST, which allows processing implicit arrays of disordered data and, on this basis, to extract new knowledge.

3. MATERIALS AND METHODS

Let's first consider RST basic concepts, which are necessary for presentation knowledge about cases, forming the CB and subsequent searching of cases (for complete information on RST, see [8, 9, 10]).

The rough set approach is considered as a concept and theoretical basis of reasoning about knowledge, when it is not exact (inexact or rough knowledge). This theory is based on the fact that knowledge is deeply embedded in the ability of people to classify subjects, phenomena, objects, situations, etc. In other words, knowledge is based on the ability to classify the objects. Therefore, knowledge in RST is necessarily associated with a multitude of patterns of classification, called the universe of discourse. In fact, knowledge consists of a family of different samples of element classification that make up the domain of interest to us.

Let $U \neq \emptyset$ be a finite set of considered objects (universe). Each subset $X_i \subseteq U$ of the universe is called a category in U and any family of subsets of the universe U is considered as abstract knowledge about U (the empty set \emptyset is also a category). RST is based on concepts which form a partition (classification) of the certain universe U , i.e. on gaining of the family $C = \{X_1, X_2, \dots, X_n\}$, such that $X_i \subseteq U$, $X_i \neq \emptyset$, $X_i \cap X_j = \emptyset$ for $i \neq j$, ($i, j = 1, \dots, n$) and $\bigcup X_i = U$.

Such family is called a knowledge base on U , which represents a set of basic aspects of classification (color, temperature, etc.) [10].

It is proposed in RST to perform the classification procedures using equivalence relations, which are simpler when dealing with them than with known decisive rules [8, 9].

If R denotes the equivalence relation then the relation $IND(R)$ denotes the family of all equivalence classes of R or classifications of U , and $[x]_R$ is represented an equivalence class in R containing an element $x \in U$.

Formally, a knowledge base, is seen as a relational system $K = (U, R)$, where $U \neq \emptyset$ is a finite set of objects (universe), R is a family of equivalence relations over U .

If to consider the goal set of elements $X \in U$, than the following situations can be considered with respect to $IND(R)$ classification [8, 9]:

1) The set X is the union of certain categories of $IND(R)$. In this case the set X is called R -exact.

2) The set X cannot be expressed as the union of certain categories of $IND(R)$. In this case the set X is called R -inexact or R -rough.

3) \underline{R} -lower approximation of rough set X is a subset of all elements of U , which can be certainly classified as belonging to the goal set X :

$$\underline{RX} = \{x \in U : [x]_R \subseteq X\}, \text{ or } x \in \underline{RX}, \quad (2)$$

if and only if $[x]_R \subseteq X$.

\underline{R} -lower approximation of X is also defined as R -positive region of X :

$$POS_R(X) = \underline{RX}. \quad (3)$$

4) \overline{R} -upper approximation of rough set X is called the subset of all elements of U , which can be possibly classified as belonging to the goal set X :

$$\overline{RX} = \{x \in U : [x]_R \cap X \neq \emptyset\}, \text{ or } x \in \overline{RX}, \quad (4)$$

if and only if $[x]_R \cap X \neq \emptyset$.

5) Negative region of X is a subset of all elements of universe U , which certainly do not belong to the set X :

$$NEG_R(X) = U - \overline{RX}. \quad (5)$$

6) The boundary region of X is a subset of all elements of U , which belong to \overline{R} -upper approximation and do not belong to \underline{R} -lower approximation of X :

$$BN_R(X) = \overline{RX} - \underline{RX}. \quad (6)$$

Let, for example [8], there be a knowledge base $K=(U, R)$, where $U=\{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}\}$, is the universe of elements x_i ; R is the equivalence relation upon which the following equivalence classes (categories) were selected on U : $U/IND(R)=\{\{x_1, x_2\}, \{x_3, x_7, x_{10}\}, \{x_4\}, \{x_5\}, \{x_6\}, \{x_8\}, \{x_9\}\}$.

Let goal sets $X_1=\{x_1, x_2, x_4, x_5\}$ and $X_2=\{x_1, x_2, x_3, x_4\}$ be given. Let us compute the characteristics considered above for sets X_1 and X_2 .

Thus, we have:

$$\underline{RX}_1 = \{x_1, x_2\} \cup \{x_4\} \cup \{x_5\} = \{x_1, x_2, x_4, x_5\};$$

$$\overline{RX}_1 = \{x_1, x_2, x_4, x_5\};$$

$$NEG_R(X_1) = \{x_3, x_6, x_7, x_8, x_9, x_{10}\};$$

$$BN_R(X_1) = \emptyset;$$

$$\underline{RX}_2 = \{x_1, x_2\} \cup \{x_4\} = \{x_1, x_2, x_4\};$$

$$\overline{RX}_2 = \{x_1, x_2, x_3, x_4, x_7, x_{10}\};$$

$$NEG_R(X_2) = \{x_5, x_6, x_8, x_9\};$$

$$BN_R(X_2) = \{x_3, x_7, x_{10}\}.$$

RST assumes that rough sets simulate the uncertainty regarding the belonging of some elements of the universe to a given goal set. Therefore, for evaluating the degree of such uncertainty, an assessment of approximation accuracy has been introduced [8, 9, 10]:

$$\alpha_R(X) = \frac{\text{card } \underline{RX}}{\text{card } \overline{RX}}, \quad X \neq \emptyset, \quad \alpha_R(X) \in [0,1]. \quad (7)$$

This assessment defines the degree of knowledge completeness.

If $BN_R = \emptyset$, i.e. $\overline{RX} = \underline{RX}$, then $\alpha_R(X)=1$ and the set X is considered as R -definable. If $BN_R \neq \emptyset$, then $\text{card } \overline{RX} > \text{card } \underline{RX}$ and $\alpha_R(X) < 1$. Such values of $\alpha_R(X)$ characterize R -undefinable or rough sets.

As an alternative measure, in RST a roughness measure of the goal set X has been introduced:

$$\rho_R(X) = 1 - \alpha_R(X). \quad (8)$$

The roughness assessment, in contrast to the accuracy assessment, characterizes the degree of incompleteness of existing knowledge.

4 EXPERIMENTS

Let's consider an example of using RST for implementation of CBR method in terms of presenting knowledge about cases, the CB creation, and also adapting and searching cases in the CB. To solve the first problem, RST uses a data table whose rows correspond to elements (cases), and columns correspond to features (attributes) of these cases.

A cell located at the intersection of i -th row and j -th column contains the value of the j -th characteristic (attribute, feature) for the i -th element. This makes it possible to obtain a simple form of the initial data representation (Table. 1).

Table 1 – The universe of cases and the values of their attributes

The universe of cases	Values of case attributes				
	a_1	a_2	a_3	a_4	a_5
x_1	1	2	0	1	2
x_2	1	2	0	1	2
x_3	2	0	0	0	1
x_4	0	0	1	1	1
x_5	2	1	0	2	1
x_6	0	0	1	1	1
x_7	2	0	0	0	1
x_8	0	1	2	1	0
x_9	2	1	0	2	1
x_{10}	2	0	0	0	1

We denote a universe as $U = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}\}$ and a set of attributes as $A = \{a_1, a_2, a_3, a_4, a_5\}$.

Using the equivalence relation, let us divide the set U into subsets of attributes like $\{a_1\}$, $\{a_1, a_2\}$, $\{a_1, a_2, a_3\}$, $\{a_1, a_2, a_3, a_4\}$, $\{a_1, a_2, a_3, a_4, a_5\}$.

So we have:

$$U/IND(R_1)\{a_1\} = \{\{x_1, x_2\}, \{x_3, x_5, x_7, x_9, x_{10}\}, \{x_4, x_6, x_8\}\};$$

$$U/IND(R_2)\{a_1, a_2\} = \{\{x_1, x_2\}, \{x_3, x_7, x_{10}\}, \{x_4, x_6\}, \{x_5, x_9\}, \{x_8\}\};$$

$$U/IND(R_3)\{a_1, a_2, a_3\} = \{\{x_1, x_2\}, \{x_3, x_7, x_{10}\}, \{x_4, x_6\}, \{x_5, x_9\}, \{x_8\}\};$$

$$U/IND(R_4)\{a_1, a_2, a_3, a_4\} = \{\{x_1, x_2\}, \{x_3, x_7, x_{10}\}, \{x_4, x_6\}, \{x_5, x_9\}, \{x_8\}\};$$

$$U/IND(R_5)\{a_1, a_2, a_3, a_4, a_5\} = \{\{x_1, x_2\}, \{x_3, x_7, x_{10}\}, \{x_4, x_6\}, \{x_5, x_9\}, \{x_8\}\}.$$

This allows us to write an expression for the CB for this example:

$$CB = (U, E_1, E_2, E_3, E_4, E_5),$$

where: $E_1 = \{x_1, x_2\}$, $E_2 = \{x_3, x_7, x_{10}\}$, $E_3 = \{x_4, x_6\}$, $E_4 = \{x_5, x_9\}$, $E_5 = \{x_8\}$ are the family of equivalence relations over U (elements in each class are considered indistinguishable).

Let $X_1 = \{x_4, x_6, x_8\}$, $X_1 \subseteq U$ and $X_2 = \{x_1, x_2, x_3, x_5, x_9\}$, $X_2 \subseteq U$, be goal subsets which need to be checked against selected classes of cases.

The first set X_1 can be uniquely represented as the union of classes E_3 and E_5 , i.e. $X_1 = E_3 \cup E_5 = \{x_4, x_6\} \cup \{x_8\} = \{x_4, x_6, x_8\}$. Therefore, X_1 is R -exact set.

The set X_2 includes classes E_1 , E_4 and one element from the class E_2 . Therefore, X_2 is R -rough set.

Using expressions (2)–(6) let's represent the following approximations of goal sets X_1 and X_2 in the following form:

$$\underline{R}X_1 = \overline{R}X_1 = POS_R(X_1) = \{x_4, x_6, x_8\};$$

$$POS_R(X_1) = \underline{R}X_1 = \{x_4, x_6, x_8\};$$

$$NEG_R(X_1) = U - \overline{R}(X_1) = \{x_1, x_2, x_3, x_5, x_7, x_9, x_{10}\};$$

$$BN_R(X_1) = \overline{R}X_1 - \underline{R}X_1 = \emptyset;$$

$$\overline{R}X_2 = \{x_1, x_2, x_3, x_5, x_7, x_9, x_{10}\};$$

$$POS_R(X_2) = \underline{R}X_2 = \{x_1, x_2, x_5, x_9\};$$

$$NEG_R(X_2) = U - \overline{R}(X_2) = \{x_4, x_6, x_8\};$$

$$BN_R(X_2) = \overline{R}X_2 - \underline{R}X_2 = \{x_3, x_7, x_{10}\}.$$

Based on the above calculations, we can confidently state that the elements $\{x_4, x_6, x_8\}$ completely belong to the goal subset X_1 and the elements $\{x_1, x_2, x_5, x_9\}$ completely belong to the goal subset X_2 .

R -boundary region of a goal subset X_2 contains three elements $BN_R(X_2) = \{x_3, x_7, x_{10}\}$. These elements can potentially belong to the goal subset X_2 .

Despite the fact that the element x_3 is specified as belonging to the goal subset X_2 ($X_2 = \{x_1, x_2, x_3, x_5, x_9\}$), element x_3 cannot be uniquely classified as element from the class X_2 , because it belongs to the boundary region of X_2 . We can only talk about its potential membership to the goal subset X_2 .

The elements x_7 and x_{10} were not assigned to any goal subset, however they, like the element x_3 , belong to the same equivalence class E_2 .

Recall that elements, which belong to the same equivalence category are considered indistinguishable, that is why the elements x_7 and x_{10} , like the element x_3 , were included in the R -upper approximation of the goal subset X_2 .

Let $X_7 = \{x_7, x_{10}\}$, $X_7 \subseteq U$. Using expressions (2)–(6) let's represent the following approximations of goal set X_7 in the following form:

$$POS_R(X_7) = \underline{R}X_7 = \emptyset;$$

$$\overline{R}X_7 = E_2 = \{x_3, x_7, x_{10}\};$$

$$NEG_R(X_7) = U - \overline{R}(X_7) = \{x_1, x_2, x_4, x_5, x_6, x_8, x_9\};$$

$$BN_R(X_7) = \overline{R}X_7 - \underline{R}X_7 = \{x_3, x_7, x_{10}\}.$$

As can be seen, from the above calculations, it is fundamentally impossible to conclude that elements x_7 and x_{10} belong to one of the classes using the information on classification attributes, because $\underline{R}X_7 = \emptyset$.

The considered elements require an increase in a priori knowledge regarding their belonging to given set of classes.

Using expressions (7) and (8) let us calculate assessments for considered goal sets $X_1 = \{x_4, x_6, x_8\}$ and $X_2 = \{x_1, x_2, x_3, x_5, x_9\}$: $\alpha_R(X_1) = 3 / 3 = 1$; $\rho_R(X_1) = 0$; $\alpha_R(X_2) = 4 / 7$; $\rho_R(X_2) = 1 - 4 / 7 = 3 / 7$.

It should be noted that elements with unknown membership to classes and belonging to the R -boundary region of classifications performed on the basis of all other elements, do not affect the classification of the original universe.

5 RESULTS

As can be seen from the above expressions the goal set X_1 completely belongs to the union of classes E_3 and E_5 ($\alpha_R = 1$). The set X_2 cannot be uniquely classified because $\rho_R(X_2) = 3 / 7$ (elements $\{x_3, x_7, x_{10}\}$).

Taking into account the considered situation, RS theory proposes next strict rules for the classification of goal sets, which can characterize cases:

if $x_i \in POS_R(X_j)$, then $x_i \in X_j$;

if $x_i \in NEG_R(X_j)$, then $x_i \notin X_j$;

if $x_i \in BN_R(X_j)$, then $x_i \in X_j$.

The above records are interpreted as follows [10]:

– if x_i belongs to the R -positive region, then it is identified with complete certainty as belonging to this class;

– if x_i belongs to the R -negative region of a particular class, it is identified with complete certainty as not belonging to this class;

– if x_i belongs to the R -boundary region of a particular class, it is impossible to say anything definite about the belonging or not belonging of the case to a particular class.

It follows that for a confident search of cases in the CB by procedure of their classification only the first decision rule can be used.

6 DISCUSSION

The proposed approach, aimed at modifying CBR method using RST, allows processing of arrays of unordered (rough) data, and on the basis of such processing to extract new knowledge. The rough set approach is based on conclusion that knowledge is reflected in ability to classify objects and is associated with a plurality of classification samples. The main feature of RST is that it uses a kind of “inaccurate” (“inexact”) classification, which in practice can look more real than an exact classification, for which it is often impossible to establish a classification category.

CONCLUSIONS

A modification of CBR method by using RST has been considered in this paper. In order to solve problems underlying this method, it is proposed to apply the following procedures of the RST: representation, structuring and case base construction, which is performed by means of a relational table and forming a family of equivalence relations (classes); for adapting cases and their search in the case base special approximations and corresponding classification rules are used.

The scientific novelty of obtained results consists in the fact that for processing and presentation of case (precedent) knowledge it is offered to use the mathematical apparatus of the Rough Set Theory. The fundamental concept of RST lies in fact that knowledge is reflected in the separation (classification) of relevant elements. Unlike existing approaches, the RST allows to correctly handle data that are rough in the sense of unordered and inaccuracy. The proposed modification of the CBR method allows solving the problems of presentation and structuring of knowledge about cases (precedents), their adaptation and subsequent search in a case base formed in the conditions of uncertainty and the incompleteness of expert information about cases.

The practical significance of obtained results indicate the possibility of forming a set of decision rules (classifier) for classification of new data samples that are not part of the original training universe. That enables to perform the description and presentation of knowledge by highlighting their properties and attributes, and creates the basis for structuring and analytical processing of inaccurate knowledge in the artificial intelligence systems design.

Prospects for further research are to develop methods to improve a classification accuracy of knowledge under incompleteness of expert information.

ACKNOWLEDGEMENTS

The work is supported by the state budget scientific research project of Petro Mohyla Black Sea National University “Development of information and communication decision support technologies for strategic decision-making with multiple criteria and uncertainty for military-civilian use” (state registration number 0117U007144).

REFERENCES

1. Varshavskij P. R., Ereemeev A. P. Modelirovanie rassuzhdenij na osnove precedentov v intellektual'nyx sistemax podderzhki prinyatiya reshenij, *Iskusstvennyj intellekt i prinyatie reshenij*, 2009, No. 2, pp. 45–57.
2. Varshavskij P. R., Ereemeev A. P. Metody pravdopodobnyx rassuzhdenij na osnove analogij i precedentov dlya intellektual'nyx sistem podderzhki prinyatiya reshenij, *Novosti iskusstvennogo intellekta*, 2006, No. 3, pp. 39–62.
3. Eremenko T. K., Pilipenko Yu. G. Ispol'zovanie CBR-podxoda dlya baz znanij situacionnyx centrov, *Systemy pidtrymky pryiniattia rishen. Teoriia i praktyka*, 2010, pp. 151–153.
4. Klymchuk S. A. Primenenie precedentov dlya diagnostiki kranov mostovogo tipa, *Systemni doslidzhennia ta informatsiini tekhnologii*, 2012, No. 4, pp. 17–22.
5. Ulshyn V. O., Klymchuk S. O. Modeli predstavlenia znan ekspertnykh system pidtrymky pryiniattia rishen pry diagnostuvanni / V. O. Ulshyn, // *Visnyk SNU im. V. Dalia*. – 2009. – Vol. 135, № 5. – S. 21–25.
6. Nefedov L. I., Fil' N. Yu., Gubin Yu. L. Metod poiska precedentov proektov likvidacii chrezvyhajnyx prirodnyx situacij na magistral'nyx avtomobil'nyx dorogax, *Vostochno-Evropejskij zhurnal peredovyx texnologij*, 2010, No. 1, pp. 50–52.
7. Haritonov Yu. N. Upravlenie proektami rekonstrukcii na osnove artefaktnyx platform, *Aviacionno-kosmicheskaya tekhnika i tekhnologiya*, 2008, No. 8, pp. 189–192.
8. Pawlak Z. *Rough Sets Theoretical Aspects of Reasoning about Data*. Boston/London, Academic Publishers, 1991, 229 p.
9. Uzhga-Rebrov O. Osobennosti predstavleniya znanij v teorii grubyx mnozhestv, *Environment. Technology. Resources: 7th International Scientific and Practical Conference, Rezekne, 25–27 June, 2009: proceedings*. Latvia, Izdevnieciba, 2009, Vol. 2, P. 169–176.
10. Uzhga-Rebrov O. I. Upravlenie neopredelennostyami. *Sovremennye neveroyatnostnye metody*. Latvia, RA Izdevnieciba, 2010, Vol. 3, 560 p.

Received 21.05.2018.
Accepted 14.06.2018.

МОДИФІКОВАНИЙ МЕТОД МІРКУВАНЬ ЗА ПРЕЦЕДЕНТАМИ НА ОСНОВІ ТЕОРІЇ ГРУБИХ МНОЖИН

Коваленко І. І. – д-р техн. наук, професор, професор кафедри інженерії програмного забезпечення Чорноморського національного університету імені Петра Могили, Миколаїв, Україна.

Швед А. В. – канд. техн. наук, доцент кафедри інженерії програмного забезпечення Чорноморського національного університету імені Петра Могили, Миколаїв, Україна.

Коваль Н. В. – аспірант кафедри інженерії програмного забезпечення Чорноморського національного університету імені Петра Могили, Миколаїв, Україна.

АНОТАЦІЯ

Актуальність. Основним елементом систем штучного інтелекту є бази знань, які формуються на основі двох загальноприйнятих підходів: об'єктно-орієнтованого та об'єктно-структурного підходів. При цьому основною операцією, яка реалізується в обох підходах, є структуризація знань за допомогою їх упорядкування та класифікації, типізації виділених класів. Досить часто виникають ситуації, коли дані або знання не є точними і неможливо виконати точну класифікацію. Зазначені особливості обумовлюють необхідність розробки нових підходів спрямованих на вирішення завдань вилучення знань з великих масивів невпорядкованих (необроблених, грубих) даних, структуривання, подання та аналітичної обробки неточних знань при автоматизованій побудові бази знань.

Метод. В роботі запропоновано підхід, спрямований на модифікацію методу міркувань за прецедентами на основі математичного апарату теорії грубих множин. Запропонований підхід дозволяє виконувати розбиття множини прецедентів для встановлення ступеня їх приналежності до заданих цільових класів на основі операції верхньої та нижньої апроксимації цільових класів, враховуючи відносну важливість класифікаційних атрибутів та виділені класи еквівалентності, з метою їх подальшої адаптації та пошуку в базі прецедентів.

Результати. Запропонована модифікація методу міркувань за прецедентами дозволяє добувати знання про прецеденти з масивів невпорядкованих даних, з метою формування бази прецедентів, та обробляти суперечливу інформацію про прецеденти, у випадку коли при однакових значеннях атрибутів, прецеденти відносяться до різних класів, та неповну інформацію про прецеденти, якщо значення деяких атрибутів або інформація про приналежність прецедентів до заданого класу відсутня або недостовірна.

Висновки. Запропонований підхід представлення знань про прецеденти, їх адаптацію та подальший пошук в базі прецедентів, сформованих в умовах невизначеності та наявності неточних, необроблених, суперечливих вихідних даних складає теоретичну підставу для побудови інтелектуальних систем підтримки прийняття рішень.

КЛЮЧОВІ СЛОВА: теорія грубих множин, метод міркувань за прецедентами, база знань, база прецедентів, класифікація.

МОДИФИЦИРОВАННЫЙ МЕТОД РАССУЖДЕНИЙ ПО ПРЕЦЕДЕНТАМ НА ОСНОВЕ ТЕОРИИ ГРУБЫХ МНОЖЕСТВ

Коваленко И. И. – д-р техн. наук, профессор, профессор кафедры инженерии программного обеспечения Черноморского национального университета имени Петра Могили, Николаев, Украина.

Швед А. В. – канд. техн. наук, доцент кафедры инженерии программного обеспечения Черноморского национального университета имени Петра Могили, Николаев, Украина.

Коваль Н. В. – аспирант кафедры инженерии программного обеспечения Черноморского национального университета имени Петра Могили, Николаев, Украина.

АННОТАЦИЯ

Актуальность. Основным элементом систем искусственного интеллекта являются базы знаний, которые формируются на основе двух общепринятых подходов: объектно-ориентированный подход и объектно-структурный подход. При этом основной операцией, которая реализуется в обоих подходах, является структуризация знаний посредством их упорядочения и классификации, типизации выделенных классов. Достаточно часто возникают ситуации, когда данные или знания не являются точными и невозможно выполнить их точную классификацию. Указанные особенности обуславливают необходимость разработки новых подходов направленных на решение задач извлечения знаний из больших массивов неупорядоченных (необработанных, грубых) данных, структурирования и аналитической обработки неточных знаний при автоматизированном построении баз знаний.

Метод. В работе предложен подход, направленный на модификацию метода рассуждений по прецедентам на основе математического аппарата теории грубых множеств. Предложенный подход позволяет производить разбиение прецедентов для установления степени их принадлежности к заданным целевым классам, используя операции верхних и нижних аппроксимаций, учитывая относительную важность классификационных атрибутов и выделенные классы эквивалентности.

Результаты. Предложенная модификация метода рассуждений по прецедентам позволяет извлекать знания о прецедентах из массивов неупорядоченных данных, с целью формирования базы прецедентов, и обрабатывать противоречивую (при одних и тех же значениях классификационных атрибутов, прецеденты относятся к различным

классам) и неполную (значения некоторых атрибутов или информация о принадлежности прецедента к целевому классу отсутствует) информацию о прецедентах.

Выводы. Предложенный подход представления знаний о прецедентах, их адаптации и последующего поиска в базе прецедентов, сформированных в условиях неопределенности и наличия неточных, необработанных, противоречивых исходных данных составляет теоретическое основание для построения интеллектуальных систем поддержки принятия решений.

КЛЮЧЕВЫЕ СЛОВА: теория грубых множеств, метод рассуждений по прецедентам, база знаний, база прецедентов, классификация.

ЛІТЕРАТУРА / LITERATURA

1. Варшавский П. Р. Моделирование рассуждений на основе прецедентов в интеллектуальных системах поддержки принятия решений / П. Р. Варшавский, А. П. Еремеев // Искусственный интеллект и принятие решений. – 2009. – № 2. – С. 45–57.
2. Варшавский П. Р. Методы правдоподобных рассуждений на основе аналогий и прецедентов для интеллектуальных систем поддержки принятия решений / П. Р. Варшавский, А. П. Еремеев // Новости искусственного интеллекта. – 2006. – № 3. – С. 39–62.
3. Еременко Т. К. Использование СВР-подхода для баз знаний ситуационных центров / Т. К. Еременко, Ю. Г. Пилипенко // Системи підтримки прийняття рішень. Теорія і практика. – 2010. – С. 151–153.
4. Климчук С. А. Применение прецедентов для диагностики кранов мостового типа / С. А. Климчук // Системні дослідження та інформаційні технології. – 2012. – № 4. – С. 17–22.
5. Ульшин В. О. Моделі представлення знань експертних систем підтримки прийняття рішень при діагностуванні / В. О. Ульшин, С. О. Климчук // Вісник СХУ ім. В. Даля. – 2009. – Том 135, № 5. – С. 21–25.
6. Нефедов Л.И. Метод поиска прецедентов проектов ликвидации чрезвычайных природных ситуаций на магистральных автомобильных дорогах / Л.И. Нефедов, Н. Ю. Филь, Ю. Л. Губин // Восточно-Европейский журнал передовых технологий. – 2010. – № 1. – С. 50–52.
7. Харитонов Ю. Н. Управление проектами реконструкции на основе артефактных платформ / Ю. Н. Харитонов // Авиационно-космическая техника и технология. – 2008. – № 8. – С. 189–192.
8. Pawlak Z. Rough Sets Theoretical Aspects of Reasoning about Data / Z. Pawlak. – Boston/London: Academic Publishers, 1991. – 229 p
9. Ужга-Ребров О. Особенности представления знаний в теории грубых множеств / О. Ужга-Ребров // Environment. Technology. Resources: 7th International Scientific and Practical Conference, Rezekne, 25–27 June, 2009: proceedings. – Latvia : Izdevnieciba, 2009. – Vol. 2 – P. 169–176.
10. Ужга-Ребров О. И. Управление неопределенностями. Современные невероятностные методы / О. И. Ужга-Ребров. – Латвия: RA Izdevnieciba, 2010. – Ч. 3. – 560 с.