

THE EXPERT KNOWLEDGE COLLECTION METHODOLOGY IN THE DECISION SUPPORT SYSTEM

Anna MICHALCZYK*, Tadeusz KRUPA**

Faculty of Management
Warsaw University of Technology, 02-524 Warszawa, Poland

*email: aniam@twp.olsztyn.pl

**email: t.krupa@wz.pw.edu.pl

Abstract: The purpose of this article is to present and analyse issues connected with the expert knowledge collection methodology in the decision support systems (DSS). The considerations concentrate on a conception of building an information system, based on an application of case-based reasoning method and reasoning based on approximate knowledge. The expert's knowledge is systematically collected in a case base. A mechanism of classical CBR and a logic model of the case base were described. It was assumed that the cases compared with regard to similarity are elements of tolerance space what considerably accelerates the retrieval of satisfying solutions. A local and global measure of case similarity is developed. The method can be used in complex tasks of image identification.

Keywords: Expert system, case-based reasoning (CBR), approximate knowledge, object-attribute-value, tolerance space, CBR mechanism, expert inference, case, case base, similarity of phenomena, matrix of similarity, CBR cycle.

1 Introduction

The expert knowledge – is the experts' knowledge (hence the term: "expert system"). This knowledge requires a continual verification within the scope of its correctness and effectiveness. Hence there is a need of quality evaluation and its social usefulness.

The experts' knowledge working with DSS is an example of an immeasurable factor whose value becomes a real value in a situation when the expert gives up the job – hence there is a necessity to collect (codify) the experts' knowledge in a special DSS knowledge base, whose element is a case base. The case base collects the experience in the form of cases, understood as problems and their solutions, without instructions related to the rules on the basis of which these decisions were made. The experience is a specific knowledge developed when solving the problems, connected with a concrete problem situation, easy to observe and describe, gained through actions and participation

in the events. Requiring the experience of the expert, it is expected that the current problems shall be solved basing on the cases that occurred in the past. It is expected that starting to solve a current problem the expert shall try to recall what solutions were applied in a similar situation in the past. If they lead to the success, surely the expert shall try to act in the same or similar way in a given situation as well. If they came to

fail earlier and undertaken actions were ineffective, this time surely the expert shall take advantage of the experience and not reuse applied methods, but replace them with other actions. A presented mechanism of using the experience corresponds with reasoning by analogy whose essence lies in the fact that a reasoning person, in the course of solving a problem, goes back in his mind to the past, recalls problems known to him, presents them and models himself upon them. The use of earlier collected experience requires not only to store it in a convenient form but to introduce automated mechanisms of using it as well. A case-based reasoning method (CBR method) – reasoning by analogy – is an artificial intelligence method which imitates a mechanism of experience use.¹

The purpose of this article is to present and analyse selected issues connected with the expert knowledge collection methodology in the decision support systems. The considerations concentrate mainly on presenting a conception of building an information system with a case base, based on a case-based reasoning method and assumption of making use of approximate

¹ A CBR conception was presented in the study by Riesbeck C., Schank R., *Inside Case-Based Reasoning*, Lawrence Erlbaum, 1989, later developed by Kolodner J., *Case-Based Reasoning*, Morgan Kaufman Publishers Inc., San Francisco 1993. The information based on Maher M., Balachandran M., Zhang D., *Case-based reasoning in design*, Lawrence Erlbaum Associates, 1995, p. 2 [on:] <http://books.google.com> (15.01.2010).

knowledge by a reasoning person². A representation conception of the expert's knowledge in the form of case base is presented in the article. A mechanism of classical CBR and an exemplary logic model of the case base were discussed. Then it was assumed that the cases compared with regard to similarity are elements of tolerance space what allows to realize an introduced assumption. The tolerance spaces as a special type of approximation space were described among other in the study [5]. The elements of these spaces can be compared with each other with regard to similarity. It is assumed that apart from a set of objects (elements), the additional knowledge about similarity within the framework of their features (properties) is possessed as well, however, it is assumed that the similarity occurred in complex structures is based on the similarity occurred in simpler, elementary sets of values. The assumption that the cases are elements of tolerance space affects the CBR mechanism what is presented in a chapter describing an information system model. There is a presentation of expert inference as an example of system operation of a given model in the article as well. A summary concludes the article.

2 The expert's knowledge recorded in the case base

A CBR method or method of the system with a case base is used in making prognostic decisions based on the experience recorded in the case base. It found an application for solving many issues, such as: diagnosing, classification, forecasting, evaluation, interpretation, planning, designing, teaching, knowledge, and experience management [2]. It can be use among other to diagnose events or schedule operations. An application area of the systems with a case base includes these domains which meet the criteria such as: regularity (i.e. phenomena are predictable, execution of the same action for a subsequent time, in the same or similar situations it leads to the same or similar results), regularity, i.e. repeatability, similarity of phenomena, continuity of modelled reality (that is little changes in a modelled domain result in little changes in a way of solving old problems).

The CBR method is based on observation of the expert's reasoning who goes back in his mind to the past,

² For the needs of the study it is assumed based on [5], that approximate knowledge is the knowledge arisen on the basis of approximate data encumbered with an error.

recalls problems known to him, presents them and models himself upon their solutions in the course of problem solving process. Therefore CBR takes advantage of specific knowledge included in situations, experienced in the past, called cases. According to this approach, to put it most generally, a case is a pair: <problem, solution>³. A problem as well as a solution posses features (attributes) which can be described by the data of various types of values: e.g. by numbers, symbols, text, sets of values, multimedia, etc. The cases are independent, they are not rules, they are a record of real events. Therefore the case base collects experience in the form of cases, understood as problems and their solutions, without instructions related to the rules on the basis of which the decisions were made. This property distinguishes a system with the case base from an expert system where the knowledge is expressed by means of the rules.

In the systems based on rule-based reasoning, the knowledge is represented by means of facts and rules of if-then type. Reasoning is of cause-effect character. The creation of rule-based expert system requires the knowledge "how" to solve a problem. Therefore the purpose of the construction of the rules is generalization and structurization of the expert's knowledge. The drawbacks of rule-based representation appear in the case when the domain expert is not able to express every possible situation then the built knowledge base can include: gaps (they are constituted by unforeseen situations), discrepancies (they can be removed but it is not easy) and effects of personal opinions of the experts (lack of objectivity). The CBR method is an alternative to rule-based reasoning because – as Riesbeck and Schank state [12]: people-experts are not systems of rules but libraries of experiences.

The knowledge representation by means of cases is the most natural form because it is easier for the domain expert to express his opinion on individual cases (examples), on each severally. It is explained by the fact that the expert gains experiences through the contact and handling of concrete situation in practice, he is engaged in solving concrete problems. It should be noticed that the experts as well as all people use analo-

³ Bergmann R., Kolodner J., Plaza E. – *Representation in case-based reasoning*, The Knowledge Engineering Review. Vol. 00:0, 1-4, Cambridge University Press, UK 2005. [on] http://www.iiaa.csic.es/People/enric/papers/Representation_in_CBR.pdf, p. 1 (18.11.2009). In the references [2, 13] it is also mentioned the third element of the case – outcome which as an optional one was omitted in the study.

gies, comparing a new, unknown case with the known cases, held in memory, with which they managed already earlier. Just as all people in everyday life, the experts make use of approximate knowledge, arisen on the basis of approximate data encumbered with an error. However it is necessary to be prepared for the fact that evaluations expressed in this way shall include some degree of uncertainty.

3 The CBR mechanism

The solution of current problem through adaptation of solutions used in the past is the essence of CBR. Therefore an idea of the method boils down to an assumption that similar problems have similar solutions. The below figure presents a graphic interpretation of the essence of CBR (see Figure 1).

The CBR mechanism is realized in the 4-phase cycle:

- retrieving a case (or cases) the most similar to the one under consideration in a case base,
- reusing a way of solution of this case (these cases) to solve a current problem,
- revising an old solution (solutions) with the object of its adjustment to the problem under consideration; possible adaptation, modification of solution,
- retaining the problem under consideration together with the used solution as a new case, experience

in order to use it later during solving new problems in the future.

Wanting to build an information system using a CBR method it is necessary to challenge many problems into which there should be included [1, 3]: definition of case within the framework of domain, selection of case representation, determination of a way of similarity measurement, selection of retrieval techniques of similar cases and solution adaptation methods.

A next problem related to the CBR is to determine a way of retrieval of cases similar to a given one. In practice, two retrieval techniques dominate [1, 7]: nearest-neighbour retrieval method which consists in reviewing a base, case by case, in order to find the most similar case and inductive retrieval method which creates a decision tree enabling to classify the cases with regard to a decision. The analysis of the references [1, 2 and 13] allows to state that in most CBR applications a nearest-neighbour retrieval method is selected.

A basic problem related to the CBR is to measure similarity between cases; the measurement depends on a way of case representation. The measurement of similarity between cases is based on the following idea: similarity of cases results from similarity occurring within the framework of their features (properties) describing a problem.

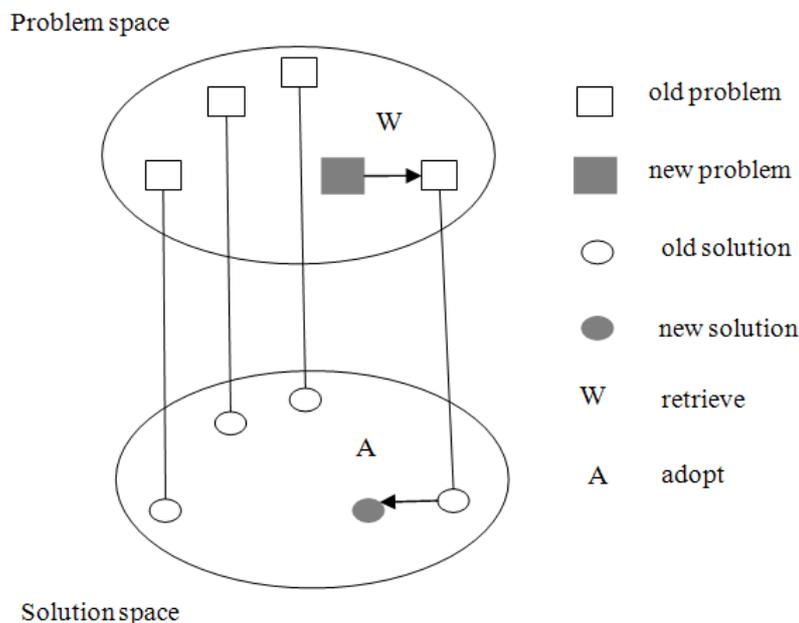


Figure 1. The CBR acc. to Leake
(source: [3])

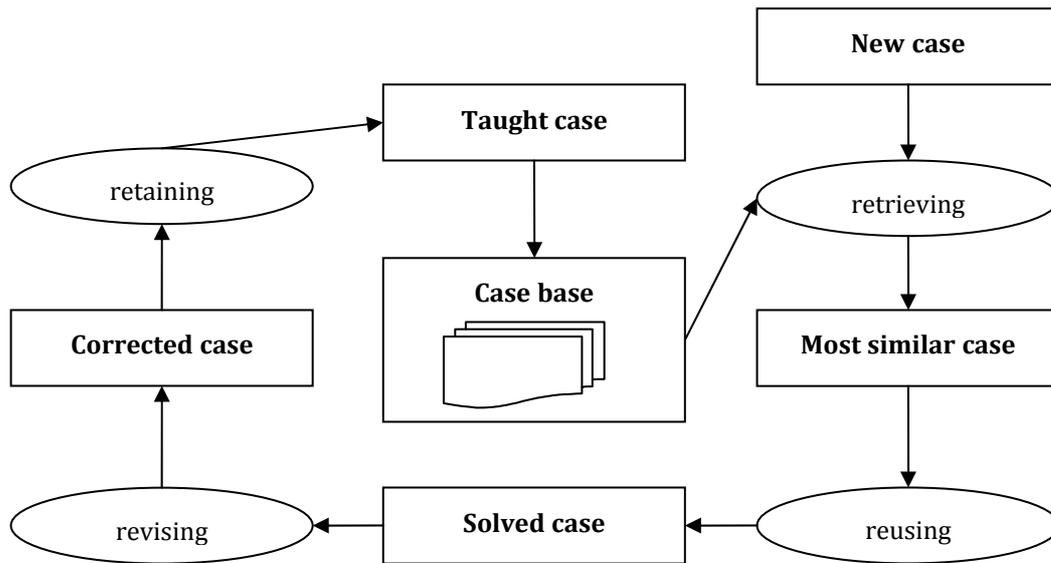


Figure 2. The CBR cycle acc. to Aamodt & Plaza, 1994
(source: self study on the basis of [2])

Additionally it is assumed that a measure of similarity of cases is a reflexive function (every case is similar to itself) and symmetric function (the following property occurs: if a case c is similar to a case c' , then c' is similar to c). A way of realization of the afore-mentioned idea of similarity measure is a structure of so called global measure, determined for the problem as a combination of so called local measures, determined for the features describing the problem, however, distinction of “power” of their impact on similarity of the case through the introduction of weights is admitted.

A next problem connected with the realization of CBR cycle is to modify, adapt a solution. As the references show, an adaptation of the solution can take place: interactively, with participation of the user who proposes the solution, or automatically, on the basis of implemented method. The simplest automatic method is to download (copy) a solution applied to similar problems, i.e. to accept the solution without its modification. A more advanced approach is to transform retrieved solutions by means of common rules or expressions.

The CBR method has many advantages but it is not deprived of drawbacks and limitations. These features which distinguish it from other techniques of automatic solving of the tasks can be included into the advantages of CBR. First of all, it does not require a good understanding of a domain from which the data come. The knowledge about the domain has not to be known, it can be incomplete, it can be gained during going

through and reading the content of cases, therefore this method allows for learning on the basis of successes and errors.

When building a CBR system, a way of “how” to solve a problem has not to be known, it is sufficient to give a solution without pointing to the rules on the basis of which it arose. This feature significantly distinguishes the CBR from the systems based on the rules. Teaching the system based on the CBR, consisting in adding the cases to the base, is automatic to some degree. Such systems process the data of any types, any complex structures, and they always give a possibility of reading on the basis of what premises a final outcome is received. The other advantages of CBR mentioned in the reference [4] include a reduction of knowledge acquisition costs, decrease of efforts made to solve a new problem, easiness of implementation of the method, relatively small maintenance cost, use of existing data, quick proposal of problem solutions, easiness of method application, simple learning through memorizing, and first of all, high acceptability by the final users.

It results from the analysis of the references [2, 13] that a very little attention is paid to the drawbacks of CBR method. In the authors’ opinion it can be assumed that in the first place the CBR drawbacks result from drawbacks and limitations of applied techniques in each of phases of CBR cycle and they can be considered as inconveniences. In the first place such inconveniences of CBR, mentioned in the reference [1] - in the authors’

opinion – can include the fact that the limitations are put on properties of the domain in which it can be used (regularity, repeatability, similarity of phenomena, continuity of modelled domain). The fact that it is necessary to select the problem features (attributes) in relation to which reasoning is carried out can be considered as a next drawback. It requires an additional knowledge of the user who carries out the reasoning. The fact that the case base requires a lot of memory can be considered as an additional difficulty as well. It can be explained in such a way that a presented classical mechanism of CBR always finds a solution therefore the fact whether a found case is more or less similar to the case under consideration depends on the size and quality of the case base. The bigger case base, the better approximation of cases. The next difficulty is a necessity of preparation, indication of similarity function adjusted to the representation of cases what can be an additional problem for the User. The classical CBR does not take making use of approximate knowledge into consideration as well what - in the authors' opinion - can be a drawback of the method.

4 The exemplary logic model of the case base

The information system created using a method of artificial intelligence, including the CBR, belongs to so called systems with a knowledge base. Three permanent elements can be distinguished in the systems of such a class [10, 11]: knowledge base, reasoning mechanism and the user's interface. In the case of CBR method application, the knowledge base is constituted by a case base and an additional knowledge about a domain, i.e. a dictionary which describes a definition of the case and its representation, similarity measures and solution adaptation rules.

The case base includes a description of various problems from the past along with the description of their solving without instructions related to the rules on the basis of which the decisions were made. It results from this determination that the case base is, in the sense of logic model of relational data base, simply a data base.

Assuming that in most CBR applications a representation based on a vector of features ("flat" model) which employs a form of structure: Object-Attribute-Value [3, 13] is selected, the case base can be defined as follows:

$$CB = \langle U, A, D \rangle \quad (1)$$

where:

U - finite set of cases,

A - non-empty, finite set of attributes (features, properties) describing a problem and for every attribute $a \in A$ we have $a: U \rightarrow U_a$, where U_a is a set of values of attribute $a \in A$,

D - non-empty, finite set of attributes describing a problem solution and for every $d \in D$ we have $d: U \rightarrow U_d$, where U_d is a set of values of attribute d.

The nearest-neighbour retrieval method as a retrieval technique of similar cases is used the most often when a representation of the case is determined in such a way. Then the measurement of similarity between two cases $c, c' \in U$ is calculated according to the formula:

$$\text{sim}(c, c') = \frac{\sum_{a \in A} w_a \text{sim}_a(a(c), a(c'))}{\sum_{a \in A} w_a} \quad (2)$$

where

A - set of attributes describing a problem and for $a \in A$ we have $a: U \rightarrow U_a$,

$\text{sim}_a: U_a \times U_a \rightarrow [0, 1]$ - local measure of similarity determined for every attribute $a \in A$,

U_a - set of values assumed for every $a \in A$,

$w_a \geq 0$ - weight of attribute $a \in A$.

It should be noticed that in the formula (2) the measures of local similarity sim_a are dependent on a type of values assumed within the framework of feature. Additionally, it is assumed that a measure of similarity of cases is a reflexive function and symmetric function, however, the weights allow to express the validity of attribute, impact of feature on similarity of cases. It is assumed that local similarities as well as weights of attributes must be known and given by the user of the system or calculated using machine training techniques. The analysis of the references [2, 13] allows to conclude that the cases of heterogeneous, mixed features – both quantitative and qualitative ones – using real, non-transformed data, can be compared with each other in this way with respect to similarity.

The executed empirical researches [9, 14] point to the problems of the users of the system with a numerical determination of local similarity. Sometimes it is significantly easier to determine a fact of similarity between the values of a given attribute then to express it numerically through giving a precise value or function (measure) according to which these value is calculated. It can be grounded by the fact that a man, in everyday

life, in a natural, intuitive way, does not calculate but he estimates, i.e. he makes use of approximate knowledge. Additionally, it can be assumed that the values of some attributes can come from a measure, therefore it is necessary to take a device error tolerance into account. The achieved results point to the demand for the extension of classical measure of similarity by measures which take making use of approximate knowledge into account. An assumption that the cases are elements of tolerance space is the attempt to solve this problem.

5 The role of tolerance space in processing the expert's knowledge

The tolerance spaces (TS) and their properties were described by Doherty, Łukaszewicz and Szałas in the article [5]. The tolerance spaces as a special kind of approximation spaces allow, in a convenient way for the user, to compare objects, whose features assume the values of various types, with each other with regard to similarity. It is assumed that apart from a set of objects, an additional knowledge about similarity within the framework of their features (properties, attributes) is known as well, however, it is assumed that the similarity occurred in complex structures is based on the similarity occurred in simpler sets of values on the basis of which they arose. It means that the tolerance spaces determined for the elementary sets (e.g. values of features) induce, determine a tolerance space for complex structures for the construction of which they were used [5]. The property is very essential because in the CBR method, the similarity of cases results from the similarity of their problems and it results from the similarity within the framework of compared features describing the problem.

Let U be a set of some objects. The similarity between the objects of the set U can be measured, i.e. it can be expressed as a value of some function, so called measure of similarity $\text{sim}: U \times U \rightarrow [0, 1]$ such one that for every $u, u' \in U$ it occurs:

- $\text{sim}(u, u') = 1$, i.e. every object is similar to each other,
- $\text{sim}(u, u') = 0$, when u and u' are not similar to each other,
- $\text{sim}(u, u')$ expresses a similarity degree of the object u' to u .

The measure of similarity increases in conjunction with the increase of similarity between objects of the set U . The objects for which a similarity degree is above

a certain threshold value can be considered as indistinguishable objects.

A concept of tolerance space (TS) is understood as a tuple $\text{TS} = \langle U, \text{sim}, p \rangle$, where U is a non-empty set of objects, $\text{sim}: U \times U \rightarrow [0, 1]$ is a measure of tolerance, p is a threshold value, however, $p \in [0, 1]$.

The measure of similarity is a tolerance measure when it is a reflexive and symmetric function, i.e. for every object $u, u' \in U$ a condition $\text{sim}(u, u') = 1$ is fulfilled and $\text{sim}(u, u') = \text{sim}(u, u')$.

A set of indistinguishable objects with the object $u \in U$, i.e. similar ones to a degree of at least p constitutes its surroundings (or neighbourhood) understood as:

$$I(u) = \{u' \in U; \text{sim}(u, u') \geq p\} \quad (3)$$

As it is given earlier, an important property of TS is the fact that the similarity of objects of complex structures results from the similarity occurring in the sets of values used for the purposes of their construction.

Let a set of objects under consideration be defined as an information table.

A concept of information table is understood as a pair $\text{IT} = \langle U, A \rangle$, where U is a non-empty, finite set of objects, A is a finite set of attributes (features, properties) and for the attribute $a \in A$ we have $a: U \rightarrow U_a$, where U_a is a set of values of attribute $a \in A$.

In the TS theory it is assumed that apart from the information table $\text{IT} = \langle U, A \rangle$ an additional knowledge about similarity occurring in the sets of values of attributes (in so called elementary sets) is possessed as well. The knowledge is expressed by means of tolerance space $\text{TS}_a = \langle U_a, \text{sim}_a, p_a \rangle$ for every attribute $a \in A$, where U_a is a set of values of attribute $a \in A$ (treated as an elementary set), sim_a means a tolerance measure determined for the elements U_a , however, p_a means a threshold value determining the surroundings. The similarity occurring in the elementary sets is subsequently transferred to the similarity of objects of information table. Then a tolerance space $\text{TS}_{\text{IS}} = \langle U, \text{sim}, p \rangle$ for the information table IT is defined through constructing a tolerance measure sim (so called global measure) on the basis of tolerance measures sim_a (so called local measures) and giving a threshold value p , e.g. by the expert or using machine training techniques.

The measures of similarity, including tolerance measures, can be defined in several ways [8, 9]. The exemplary local sim_a and global sim tolerance measures, to-

gether with a short characteristic, are presented in the further part of the study.

It should be noticed that in the CBR a retrieval of similar case in the case base boils down to comparing the problems with respect to similarity, therefore a subset CB, which with the assumed representation $CB = \langle U, A, D \rangle$ is the information table $IT = \langle U, A \rangle$, is taken into account in the phase of retrieval. Taking this fact into consideration it can be acknowledged that the tolerance measures for the information table can be measures of similarity for the cases from CB.

6 The exemplary local tolerance measures sim_a

The similarity (tolerance) of objects, being elements of elementary set (e.g. numbers, symbols, logic values) can be presented in various ways, e.g. as a matrix of similarity, affinity to surroundings and as functions of measure [8, 9]. Finally, the expert decides a way of expression of the knowledge about similarity.

- The matrix of similarity

If the similarity between objects under consideration is known and expressed by means of numerical values then these values can be stored in a so called matrix of similarity:

$$\text{MatrixSim}_a = [s_{uu'}], \text{ where } u, u' \in U_a \quad (m1)$$

and $s_{uu'} = sim_a(u, u')$ for $a \in A$

- The measures based on a distance function

If on the set of values of attribute U_a a distance function $d: U_a \times U_a \rightarrow [0, \infty]$, such one that for every $u, u' \in U_a$ it occurs $d(u, u) = 0$ and $d(u, u') = d(u', u)$ (weaker than metrics, without triangle condition) is determined, then the similarity between objects results from their proper "proximity". The smaller distance between the objects the bigger similarity between them. A selection of distance function d can depend on a kind of values of the set U_a (interval, symbolic, binary ones). Let $u, u' \in U_a$, then the measure of similarity sim_a can be expressed by the distance function according to the definition (m2):

$$sim_a(u, u') = 1 - d(u, u') / \max \{d(x, y) : x, y \in U_a\} \quad (m2)$$

A measure (m2) can be applied to these sets of values for which a calculation of maximal distance of objects is possible.

$$sim_a(u, u') = \begin{cases} 1 & \text{for } d(u, u') \leq \varepsilon; \varepsilon > 0 \\ 0 & \text{otherwise} \end{cases} \quad (m3)$$

A measure (m3) is applicable to the numerical sets when an approximation error $\pm \varepsilon$ is known. In particular, it receives the following form (m4)

$$sim_a(u, u') = \begin{cases} 1 & \text{for } u = u' \\ 0 & \text{for } u \neq u' \end{cases} \quad (m4)$$

Other exemplary tolerance measures based on the distance function are described by the formulas (m5) and (m6)

$$sim_a(u, u') = 1/(1 + d(u, u')) \quad (m5)$$

$$sim_a(u, u') = e^{-d(u, u')} \quad (m6)$$

The presented measures are applicable when a distance function d is known, adjusted to such a type of data. For example, for the numerical data, the distance can be calculated according to the formula $d(u, u') = |u - u'|$ (Euclid distance), for logic (symbolic) values, e.g. $d(u, u') = \{1 \text{ for } u = u'; 0 \text{ for } u \neq u'\}$.

- The affinity to the surroundings

If a numerical value of similarity degree is not known or a measure function is not known, or it is a problem to give its definition, then the fact of similarity between the objects can be expressed through the affinity to the object's surroundings. The knowledge can be expressed as follows.

Let

$$I_a(u) = \{u' \in U_a; sim_a(u, u') \geq p_a\}$$

mean the neighbourhood of element $u \in U_a$, then for $u, u' \in U_a$ we have:

$$sim_a(u, u') = \begin{cases} 1 & \text{for } u' \in I_a(u) \\ 0 & \text{for } u' \notin I_a(u) \end{cases} \quad (m7)$$

All presented measures sim_a are tolerance measures, i.e. they are reflexive and symmetric functions.

7 The exemplary global tolerance measures sim

The similarity (tolerance) of objects, being elements of information table (what with the assumed case representation it boils down to similarity between cases from the case base) can be presented in various ways, e.g. as a matrix of similarity, affinity to surroundings and as functions of measure based on the similarity occurring in the sets of features (elementary sets) [8, 9]. Finally, the expert decides a way of expression of the knowledge about similarity between the objects.

- The matrix of similarity

If the similarity between objects $u, u' \in U$ is known in the form of numerical values then these values can be stored in a so called matrix of similarity:

$$\text{MatrixSim} = [s_{u,u'}], \text{ where } u, u' \in U \text{ and } s_{u,u'} = \text{sim}(u, u') \quad (\text{M1})$$

- The measures based on a similarity degree within the framework of features

We assume that the similarity between the objects is not known but we have an additional knowledge about a set of values of every attribute $a \in A$ expressed by tolerance spaces $TS_a = \langle U_a, \text{sim}_a, p_a \rangle$ which induce a tolerance space $TS_{IS} = \langle U, \text{sim}, p \rangle$ for the information table. Then the function $\text{sim}: U \times U \rightarrow [0, 1]$ can be defined by means of tolerance measures sim_a of elementary sets and a threshold value p can be given by the expert or determined by means of machine training techniques. A construction of global tolerance measure on the basis of local tolerance measure can be as follows:

$$\text{sim}(u, u') = \frac{\sum_{a \in A} w_a \text{sim}_a(a(u), a(u'))}{\sum_{a \in A} w_a}, \quad (\text{M2})$$

where $w_a \in [0, 1]$

The similarity between the objects in (M2) is decided by the weight of attributes w_a and similarity degree within the framework of features (attributes), i.e. the value sim_a . This measure is a classical measure used in CBR (formula 2):

$$\text{sim}(u, u') = \left(\prod_{a \in A} \text{sim}_a(a(u), a(u')) \right)^{1/|A|} \quad (\text{M3})$$

The similarity between the objects in (M3) is decided not only by the similarity occurring within the framework of all $a \in A$, but by their accordingly big degree expressed by sim_a .

$$\text{sim}(u, u') = \max_{a \in A} (\text{sim}_a(a(u), a(u'))) \quad (\text{M4})$$

The similarity between the objects in (M4) is decided by the biggest similarity degree sim_a for $a \in A$. It means that the similarity of the objects is determined on the basis of one feature, not always the same one.

$$\text{sim}(u, u') = \min_{a \in A} (\text{sim}_a(a(u), a(u'))) \quad (\text{M5})$$

The similarity between the objects in (M5) is decided by the smallest similarity degree sim_a for $a \in A$, i.e. all features impact on the similarity.

$$\text{sim}(u, u') = \frac{\sum_{a \in A} \text{sim}_a(a(u), a(u'))}{|A|} \quad (\text{M6})$$

The similarity between the objects in (M6) is decided by an arithmetic mean of the similarities in the set

of values of attributes $a \in A$. This measure is a special case of measure (M2), assuming that the weights w_a for $a \in A$ are equal for every attribute and they receive the value $w_a = 1/|A|$.

- The measure based on the affinity to the surroundings within the framework of features

If a value of similarity degree within the framework of features is not known but the fact of “being indistinguishable” is stated, i.e. the affinity to the object’s surroundings for every feature is known, then the measure of similarity for the object can be expressed as follows.

Let sim_a mean a tolerance measure in a set of values U_a of attribute $a \in A$, p_a a threshold value, $I_a: U_a \rightarrow 2^{U_a}$ surroundings (neighbourhood) of element of this set, i.e. for $u \in U$ we have $I_a(u) = \{u' \in U_a; \text{sim}_a(u, u') \geq p_a\}$ and let $S_a: U \times U \rightarrow [0, 1]$ be a function verifying an affinity to the object y to the object’s surroundings x , where $x, y \in U_a$ for $a \in A$ described by the formula:

$$S_a(x, y) = \begin{cases} 1 & \text{for } y \in I_a(x) \\ 0 & \text{for } y \notin I_a(x) \end{cases} \quad (\text{M7})$$

then for $u, u' \in U$ we have

$$\text{sim}(u, u') = \frac{\sum_{a \in A} w_a S_a(a(u), a(u'))}{\sum w_a}, \quad (\text{M8})$$

where $w_a \in [0, 1]$ is a weight of $a \in A$

Selecting a weight of attribute w_a for $a \in A$ in (M8), the “power” of similarity within the framework of a given feature on the similarity between the objects can be expressed. The similarity between the objects is decided not only by the number of attributes for which the similarity of features occurs but also by their “quality” expressed by the weight. Selecting a threshold value p_a properly, it can be determined which attributes (features) have the biggest impact on the similarity. Assuming $w_a = 1/|A|$ for every $a \in A$, i.e. determining that the similarity within the framework of every feature is equally important, the similarity between the objects shall be decided by a number of attributes for which the indistinguishability occurs within the framework of the features.

- The affinity to the surroundings

If a numerical value of similarity degree is not known or a measure function is not known (e.g. it is a problem to give its definition, then the fact of similarity between the objects (to be more precisely: being indistinguishable) can be expressed through the affinity to the surroundings (neighbourhood). The knowledge can be expressed as follows.

Let $u \in U$ and $I(u) = \{u' \in U; \text{sim}(u, u') \geq p\}$ mean the surroundings (neighbourhood) of element u , then for $u, u' \in U$ we have (M9):

$$\text{sim}(u, u') = \begin{cases} 1 & \text{for } u' \in I(u) \\ 0 & \text{for } u' \notin I(u) \end{cases} \quad (\text{M9})$$

All presented measures are tolerance measures, i.e. they are reflexive and symmetric functions. Finally, their usefulness or purposefulness of use is decided by the expert who carries out the reasoning.

The threshold values p_a and necessary weights w_a for $a \in A$ must be determined. They are given by the user who carries out the reasoning. When the user has no knowledge on this issue, default values (threshold values equal to 1) shall be accepted or they shall be calculated automatically using machine training techniques.

The presented definitions of measures assume that for every object $u \in U$ all values for every attribute $a \in A$ are known, i.e. a problem of missing data is not taken into consideration in a proposed approach.

8 The model of information system with the case base

As it is already described, wanting to build an information system using a CBR method it is necessary to challenge many problems. The basic problems include: definition of case, selection of case representation and measurement of similarity between cases. The purpose of the article is to present a conception of the system in a generalized form, therefore a definition of the case⁴ is omitted in the presented model, it is only assumed that the case representation is based on a vector of features.

We assume that apart from the case base, an additional knowledge about occurring similarity within the framework of features of the problem under consideration, expressed by means of tolerance spaces, is known. Then this similarity is transferred to the similarity of problems and consequently to the similarity of cases. Such an assumption is a system model assumption which can be presented as follows.

The system assumptions. We have a case base $CB = \langle U, A, D \rangle$ defined according to the formula (1) and we possess an additional knowledge about similarity occurring within the framework of features of the

problem expressed by tolerance spaces $TS_a = \langle U_a, \text{sim}_a, p_a \rangle$ for every $a \in A$, then this similarity is transferred to the similarity of cases, i.e. the tolerance spaces TS_a determine a tolerance space above the case base $TS_{CB} = \langle U, \text{sim}, p \rangle$.

The introduction of the afore-mentioned assumption to the system allows to make use of surroundings I_a (formula (3)) determined for the values of features of the problem in order to determine a similarity between the cases, what can be described by means of the below-mentioned formula (formula (4), on the basis of M7 and M8):

$$\text{sim}(c, c') = \frac{\sum_{a \in A} w_a S_a(a(c), a(c'))}{\sum_{a \in A} w_a} \quad (4)$$

where:

$$S_a(x, y) = \begin{cases} 1 & \text{for } y \in I_a(x) \\ 0 & \text{for } y \notin I_a(x) \end{cases} \quad \text{for } x, y \in U_a, a \in A,$$

w_a - weight of attribute $a \in A$ and $w_a \in [0, 1]$.

$S_a: U_a \times U_a \rightarrow [0, 1]$ - determines the affinity to the surroundings I_a of the values of attribute $a \in A$, allows to make use of the fact of "being indistinguishable" within the framework of a given feature.

It should be noticed that a classical measure of similarity for the CBR determined by the formula (2) is also a tolerance measure (it is reflexive and symmetric), therefore it can be applied to the proposed approach. The presented definition of local and global tolerance measures constitute additionally an element extending a classical approach to the measurement of similarity between the cases.

The CBR_{TS} reasoning cycle. Assuming that cases from a case base CB are elements of tolerance space allows to make use of cases indistinguishable from the one under consideration, what impacts on a reasoning cycle, what we mark as CBR_{TS} (Tolerance-Based Case-Based Reasoning). In this situation, every case collected in the base determines its surroundings (neighbourhood), i.e. a set of cases indistinguishable from it, i.e. such ones whose similarity degree exceeds a certain threshold value.

Taking into account conditions which must be met by domains from a field of CBR application and the essence of CBR method in the form of statement: similar problems have similar solutions, it can be concluded that after the application of tolerance space to the CBR, the essence of CBR_{TS} method would extend the essence of CBR by the statement: indistinguishable problems

⁴ A concept of definition of case is understood as giving a number of attributes, their names, sets of assumed values.

have indistinguishable solutions, what is illustrated by Figure 3.

An additional advantage of such an approach is a possibility of quick statement of a precedent (a precedent in the meaning of untypical case, not appearing earlier, which can happen many times in the future). If a case the most similar to a new one does not belong to its surroundings, it means that the new case is a precedent. Such information – in the authors' opinion – would help the reasoning person to solve a problem because he would know that he deals with a new unknown situation, naturally, in the context of collected cases.

As it is said earlier, assuming that the cases from the CB are elements of tolerance space impacts not only on a similarity measure but also on a classical CBR cycle. It is obtained a four-phase CBR_{TS} (so called Tolerance-Based Case-Based Reasoning) cycle whose first phase of retrieving_{TS} is extended in relation to the classical phase of retrieving by a possibility of retrieving cases indistinguishable from the one under consideration (i.e. its surroundings-neighbourhood). In order In

order to retrieve the surroundings (neighbourhood) of new problem, it is proposed to use a neighbourhood-retrieval algorithm whose essence would consist in reviewing

a case base, case by case, in order to find all indistinguishable from the one under consideration, i.e. for which a property compliant with the formula (3) occurs. In relation to other CBR_{TS} phases it is assumed that they are compliant with 2, 3, 4 phases of classical CBR cycle. Taking the CBR_{TS} into account, it is a four-phase cycle including the following phases:

- retrieving a case the most similar to the one under consideration or cases indistinguishable from the one under consideration (i.e. its surroundings),
- reusing a way of solution of retrieved case (cases) to solve a new problem,
- revising an old solution (solutions) with the object of its adjustment to the problem under consideration; possible adaptation, modification of solution,
- retaining the problem under consideration together with the used solution as a new case.

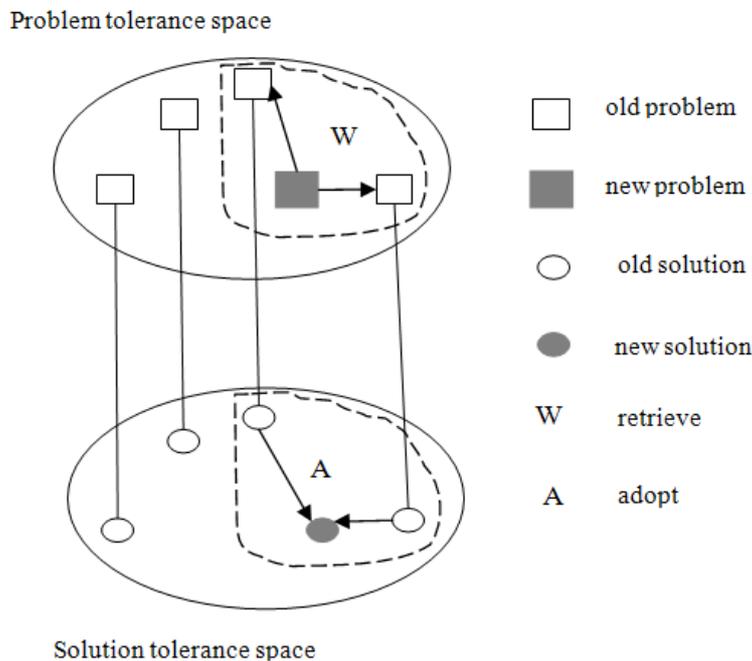
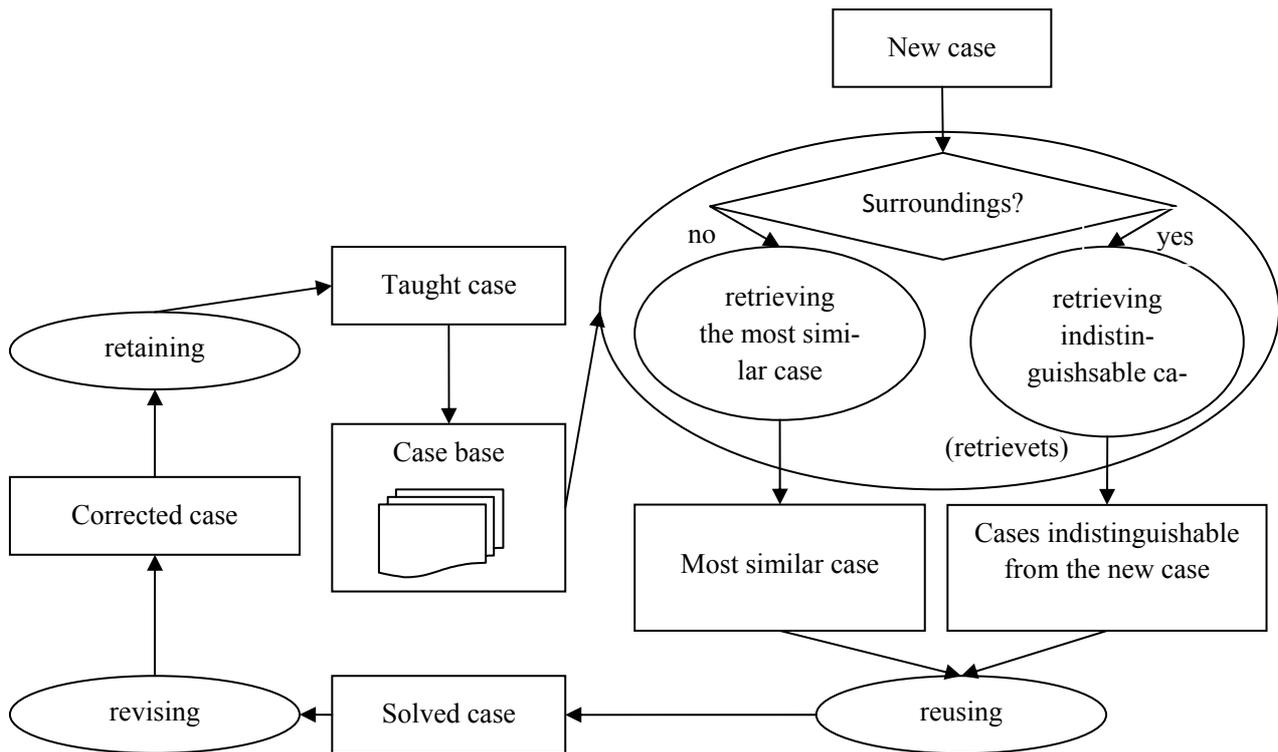


Figure 3. The essence of CBR_{TS} .
(source: self research on the basis of the CBR acc. to Leake [3])

Figure 4. The CBR_{TS} cycle.

(source: self study)

Figure 4 presents the essence of CBR_{TS} cycle. The prepared CBR_{TS} mechanism has all advantages as a classical CBR and additionally it:

- enables to make use of approximate knowledge thanks to that the user can operate with surroundings – colloquially: “a fact of being indistinguishable” without determination of similarity degree or measure,
- gives a possibility of quick statement of appearance of a precedent,
- enables to solve a new problem through adapting many solutions which were applied to the problems indistinguishable from the one under consideration,
- thanks to introduced parameterisation (use of similarity threshold) dissemination of knowledge, its repeated usage, sharing, training takes place.

A necessity of giving the thresholds p determining the surroundings can be considered as a CBR_{TS} drawback, all drawbacks of classical CBR, if a variant of retrieval of the most similar case is selected. According to this approach, a case the most similar to the new one is always retrieved, naturally, on the assumption that the case base is not empty. If the case base does not include a large number of cases, what is natural in an initial

phase of making use of CBR system, then a returned case of too small similarity degree, so that a solution connected with it is useful in a new situation, can be a result of retrieval. A possibility of return of empty surroundings as a result of retrieval which informs the user of appearance of unknown situation (precedent) can be considered as an advantage.

The system model. An information system created with using artificial intelligence tools belongs to so called systems with a knowledge base. Three permanent elements can be distinguished in the systems of this class: knowledge base, mechanism of reasoning and interface. Making assumptions about the system allows to define the information system based on a CBR method taking into account a possibility of making use of approximate knowledge in the form compliant with the formula (5):

$$\text{system} = \langle \text{CB}, \{\text{TS}_a : a \in A\}, \text{TS}_{\text{CB}}, \text{CBR}_{\text{TS}}, I \rangle \quad (5)$$

where:

CB, TS_a , TS_{CB} - create a knowledge base and are compliant with made assumptions about the system,
 CBR_{TS} - constitutes a mechanism of reasoning,
 I - the user's interface.

An exemplary operation of the system is presented in the next chapter.

A prepared model of the system presents a conception of system in the generalized form. The introduced parameterisation of the model is a tool enabling to adjust a knowledge base of the system to the user's requirements, therefore, tracking an operation of the system based on the prepared model requires:

- to design a case within the framework of domain, i.e. to give features describing the case and their characteristic in the form of assumed values U_a and U_d ,
- to organize a case base $CB = \langle U, A, D \rangle$, where U constitutes a finite set of cases, A non-empty, finite set of attributes describing a problem and for every attribute $a \in A$ we have $a: U \rightarrow U_a$, where U_a is a set of values of attribute $a \in A$ and D is a non-empty, finite set of attributes describing a solution of the problem and for every $d \in D$ of decision description we have $d: U \rightarrow U_d$, where U_d is a set of values of decision-making attribute d ,
- to give an additional knowledge about similarity within the framework of features of the problem and to express it in the form of tolerance space $TS_a = \langle U_a, sim_a, p_a \rangle$ (i.e. to determine the sets of values U_a , to point to a tolerance measure sim_a and indistinguishable threshold p_a for every feature $a \in A$),
- to give an additional knowledge about similarity of cases and to express it in a tolerance space $TS_{CB} = \langle U, sim, p \rangle$ (i.e. to determine a set of cases U , to point to a global tolerance measure sim and indistinguishable threshold p),
- to formulate a new problem,
- to select features of the problem creating an index (i.e. taking part in we reasoning) and to determine their weight w_a ,
- to execute a $CBRTS$ mechanism.

With regard to the accepted scope of the article, a model of the system (5) is presented in the basic form possible to be extended further. According to the authors' intention, the introduction of tolerance space to the CBR constitutes a starting point for the further

considerations on an application of theory of approximate set. Such a determined theme shall constitute a subject-matter of further researches.

9 The formulation of enquiries – the presentation of the expert inference

We have $CB = \langle U, A, D \rangle$ a case base, where $U = \{c0, c1, c2, c3\}$ constitutes a set of cases, $A = \{a0, a1\}$ is a set of attributes describing a problem, $D = \{d0\}$ is a set of attributes describing a decision and we possess an additional knowledge about similarity occurring in the sets of elementary attributes as in the below Table 1.

Let

- $TS_{a0} = \langle U_{a0}, sim_{a0}, p_{a0} \rangle$ be a tolerance space for the attribute $a0$, where:
 $U_{a0} = [50, 60]$,
 $sim_{a0}(u, u') = \stackrel{\text{def}}{=} 1 - d(u, u') / \max \{d(x, y) : x, y \in U_{a0}\}$,
 $p_{a0} = 0,88$ (calculated on the basis of ε),
- $TS_{a1} = \langle U_{a1}, sim_{a1}, p_{a1} \rangle$ shall be a tolerance space for the attribute $a1$, where:
 $U_{a1} = \{0, 1, 2\}$,
 $sim_{a1}(u, u') = \stackrel{\text{def}}{=} \{1 \text{ for } u' \in I(u); 0 \text{ otherwise}\}$,
 $p_{a1} = 1$ (assumed).

Then

$TS_{CB} = \langle U, sim, p \rangle$ shall be a tolerance space above the CB , where:

$$U = \{c0, c1, c2, c3\},$$

sim – variants:

- when selecting retrieval of indistinguishable ones (making use of surroundings),

$$sim(u, u') = \stackrel{\text{def}}{=} \sum_{a \in A} w_a S_a(a(u), a(u')) \quad (6)$$

$$\text{where } S_a(x, y) = \{1 \text{ for } y \in I_a(x); 0 \text{ otherwise}\}$$

- when selecting retrieval of the most similar one (classical approach),

$$sim(u, u') = \sum_{a \in A} w_a sim_a(a(u), a(u')) / \sum_{a \in A} w_a \quad (7)$$

$$w_{a0} = w_{a1} = 0,5,$$

$$p = 1 \text{ (assumed).}$$

Table 1. The case base CB and additional knowledge about similarity
(source: self study)

| | | | | | |
|--------------|------|----|-----------------|--|--|
| Case base CB | | | | Additional knowledge about similarity | |
| case | a0 | a1 | d0 | attribute a0: | |
| c0 | 50,5 | 0 | d ₀₀ | set of values $U_{a0} = [50, 60]$, | |
| c1 | 51,1 | 1 | d ₀₁ | similar values when $d(x, y) \leq \epsilon = 1,2$ | |
| c2 | 52,0 | 2 | d ₀₂ | attribute a1: | |
| c3 | 53,2 | 1 | d ₀₃ | set of values $U_{a1} = \{0, 1, 2\}$, | |
| | | | | $I(0) = \{0, 1\}; I(1) = \{0, 1, 2\}; I(2) = \{1, 2\}$ | |
| | | | | equal weight of attributes | |

Table 2. A degree of local and global similarity when retrieving indistinguishable cases, example 1, variant 1
(source: self study)

| feature a | a0 | a1 | sim(n, ci), i = 0..3 |
|--------------------------|----|----|----------------------|
| sim _a (n, c0) | 0 | 1 | 0,5 |
| sim _a (n, c1) | 0 | 1 | 0,5 |
| sim _a (n, c2) | 1 | 1 | 1,0 |
| sim _a (n, c3) | 1 | 1 | 1,0 |

The results of operation of the system described by the formula (4).

Example 1

Let $n = \langle 52, 5, 1, ? \rangle$ be a new case. The CBR_{TS} mechanism:

- Variant 1
 - Phase 1. Retrieving TS – retrieval of indistinguishable ones is selected.

Index: all features of the problem, i.e. a0, a1.

Result of the phase: surroundings of new case $I(n) = \{c2, c3\}$.

Substantiation: A degree of local and global similarity between a new case and the cases from the case base is calculated on the basis of additional knowledge about similarity between problems expressed by tolerance spaces TS_a and TS_{CB} (similarity measure based on the surroundings, formula (5)). The similarity degree is represented in the below table (Table 2).

The surroundings of a new case n include the cases which are similar at least to the degree $p = 1$ (acc. to def. TS_{CB}), therefore this condition is met by the cases c2 and c3.

- Phase 2. Reusing

A decision d_{04} can be made interactively (with participation of the reasoning person) with using the solutions d_{02}, d_{03} of cases c2, c3.

Result of the phase: d_{04} .

- Phase 3. Revising

In real conditions, this decision d_{04} is subject to revising and possible adapting and modification. It becomes a solution of new problem after the acceptance and application.

Result of the phase: new case $n = \langle 52,5, 1, d_{04} \rangle$.

- Phase 4. Retaining

The new case n being a result of the phase 3 is added to the case base as a new case $c4 = n$. The reorganization of the case base through updating a set of cases U by a new case $c4 = n$ is a result of the phase.

Result of the phase: $U = \{c0, c1, c2, c3, c4\}$, where $c4 = n = \langle 52,5, 1, d_{04} \rangle$.

- Variant 2

- Phase 1. Retrieving TS – retrieval of the most similar one is selected.

Index: all features of the problem, i.e. a0, a1.

Result of the phase: the most similar case c2.

Substantiation: A degree of local and global similarity between a new case and the cases from the case base is calculated on the basis of additional knowledge about similarity between problems expressed by tolerance spaces TS_a and TS_{CB} (classical measure of similarity, formula (7)). The similarity degree is represented in the Table 3.

The new case n is the most similar to the case c2.

Table 3. A degree of local and global similarity when retrieving indistinguishable cases, example 1, variant 2
(source: self study)

| feature a | a0 | a1 | sim(n, ci), i = 0..3 |
|--------------------------|----|----|----------------------|
| sim _a (n, c0) | 0 | 1 | 0,9 |
| sim _a (n, c1) | 0 | 1 | 0,93 |
| sim _a (n, c2) | 1 | 1 | 0,975 |
| sim _a (n, c3) | 1 | 1 | 0,965 |

Table 4. A degree of local and global similarity when retrieving indistinguishable cases, example 2, variant 1
(source: self study)

| feature a | a0 | a1 | sim(n, ci), i = 0..3 |
|--------------------------|----|----|----------------------|
| sim _a (n, c0) | 0 | 1 | 0,5 |
| sim _a (n, c1) | 0 | 1 | 0,5 |
| sim _a (n, c2) | 0 | 0 | 0,0 |
| sim _a (n, c3) | 0 | 1 | 0,5 |

- Phase 2. Reusing

A decision d_{04} can be made interactively (with participation of the reasoning person) with using the solution d_{02} of case $c2$.

Result of the phase: $d_{04} = d_{02}$.

- Phase 3. Revising

In real conditions, this decision d_{04} is subject to revising and possible adapting and modification. It becomes a solution of new problem after the acceptance and application.

Result of the phase: new case $n = \langle 52, 5, 1, d_{04} \rangle$.

- Phase 4. Retaining

The new case n being a result of the phase 3 is added to the case base as a new case $c4 = n$. The reorganization of the case base through updating a set of cases U by a new case $c4 = n$ is a result of the phase.

Result of the phase: $U = \{c0, c1, c2, c3, c4\}$, where $c4 = n = \langle 52, 5, 1, d_{04} \rangle$.

Example 2

Let $n = \langle 55, 0, 0, ? \rangle$ be a new case. The CBR_{TS} mechanism.

• Variant 1

- Phase 1. Retrieving TS – retrieval of indistinguishable ones is selected.

Index: all features of the problem, i.e. $a0, a1$.

Result of the phase: surroundings of new case $I(n) = \{ \}$ (empty).

Substantiation: A degree of local and global similarity between a new case and the cases from the case base is calculated on the basis of additional knowledge about similarity between problems expressed by tolerance spaces TS_a and TS_{CB} (measure of similarity based on the surroundings, formula (5)). The similarity degree is represented in the Table 4.

The surroundings of a new case n include the cases which are similar at least to the degree $p = 1$ (acc. to def. TS_{CB}), therefore this condition is not met by any case.

- Phase 2. Reusing

The surroundings of the new case n are an empty set, i.e. there are no cases indistinguishable from the one under consideration in the case base, therefore there is a precedent in the sense of collected cases. A decision d_{04} must be made by the reasoning person.

Result of the phase: d_{04} .

- Phase 3. Revising

In real conditions, this decision d_{04} is subject to revising. It is applied without changes or it is subject to possible adapting and modification. It becomes a solution of new problem after the acceptance and application.

Result of the phase: new case $n = \langle 55, 0, 0, d_{04} \rangle$.

Table 5. A degree of local and global similarity when retrieving indistinguishable cases, example 2, variant 2
(source: self study)

| feature a | a0 | a1 | sim(n, ci), i = 0..3 |
|--------------------------|----|----|----------------------|
| sim _a (n, c0) | 0 | 1 | 0,775 |
| sim _a (n, c1) | 0 | 1 | 0,805 |
| sim _a (n, c2) | 1 | 1 | 0,35 |
| sim _a (n, c3) | 1 | 1 | 0,91 |

- Phase 4. Retaining

The new case n being a result of the phase 3 is added to the case base as a new case c4 = n. The reorganization of the case base through updating a set of cases U by a new case c4 = n is a result of the phase.

Result of the phase: $U = \{c0, c1, c2, c3, c4\}$, where $c4 = n = \langle 55, 0, 0, d_{04} \rangle$

• Variant 2

- Phase 1. Retrieving TS – retrieval of the most similar one is selected.

Index: all features of the problem, i.e. a0, a1.

Result of the phase: the most similar case c3.

Substantiation: A degree of local and global similarity between a new case and the cases from the case base is calculated on the basis of additional knowledge about similarity between problems expressed by tolerance spaces TS_a and TS_{CB} (classical measure of similarity, formula (6)). The similarity degree is represented in the Table 5.

The new case n is the most similar to the case c3.

- Phase 2. Reusing

A decision d₀₄ can be made interactively (with participation of the reasoning person) with using the solution d₀₃ of case c3.

Result of the phase: d₀₄ = d₀₃.

- Phase 3. Revising

In real conditions, this decision d₀₄ is subject to revising and possible adapting and modification. It becomes a solution of new problem after the acceptance and application.

Result of the phase: new case n = $\langle 55, 0, 0, d_{04} \rangle$.

- Phase 4. Retaining

The new case n being a result of the phase 3 is added to the case base as a new case c4 = n. The reorganization of the case base through updating a set of cases U by a new case c4 = n is a result of the phase.

Result of the phase: $U = \{c0, c1, c2, c3, c4\}$, where $c4 = n = \langle 55, 0, 0, d_{04} \rangle$.

10 Summary

The purpose of this article was to present and analyse issues connected with the expert knowledge collection methodology in the decision support systems. The considerations concentrated mainly on the presentation of a conception of building an information system with a case base, based on a case-based reasoning method and on an assumption of making use of approximate knowledge by the reasoning person.

It was assumed that the cases are elements of tolerance space what allowed to operate with the surroundings, i.e. the cases indistinguishable from the one under consideration. It was shown that the introduction of this assumption impacted on a classical reasoning mechanism. It is received a four-phase CBR_{TS} (tolerance-based case-based reasoning) cycle being an extension of CBR method.

The case-based reasoning method and areas of its application belong to a group of subjects taken up by many researchers. Various solutions for the purpose of this method and domains, in which it can be applied successfully are still being searched. In the authors' opinion, the issues raised in the study cover this stream of researches.

11 References

- [1] Aamodt A., Plaza P. - *Case-Based Reasoning: Foundational Issues, Methodological Variation and System Approaches* [in] Artificial Intelligence Communications, No. 7 (1), 1994, pp. 39-59.
- [2] Althoff K.D., Auriol E., Barletta R., Manago M. - *A Review of Industrial Case-Based Reasoning Tools* [in] An AI Perspectives Report (ed. A. Goodall). Oxford 1995, pp. 25-39.
- [3] Bergmann R., Kolodner J., Plaza E. - *Representation in case-based reasoning* [in] Knowledge Engineering Review, No. 20 (3), Cambridge University Press, UK 2005, pp. 209-213.

- [4] *Case-Based Reasoning*. www.aaai.org/AITopics/html/casebased.html, 18.11.2009.
- [5] Doherty P., Łukaszewicz W., Szałas A. - *Tolerance Spaces and Approximative Representational Structures* [at] Proceedings of the 26th German Conference on Artificial Intelligence KI'2003, LNAI 2821, Springer-Verlag 2003, pp. 475-489.
- [6] Leake D. - *CBR in Context: The Present and Future* [in] *Case-Based Reasoning: Experiences, Lessons and Future Directions* (ed. D. Leake). AAAI Press/MITTPress 1996, pp. 3-30.
- [7] Marir F., Watson I.D. - *Case-Based Reasoning: A Review* [in] *Case-Based Reasoning: A Categorized Bibliography* (ed. F. Marir, I.D. Watson). The Knowledge Engineering Review, Vol. 9, No. 4, 1994, pp. 355-381.
- [8] Michalczyk A. - *Miary podobieństwa w hierarchicznym grupowaniu elementów przestrzeni tolerancji* [in] *Informatyka w globalnym świecie* (ed. J. Kisielnicki). PJWSTK, Warszawa 2006, pp. 357-363.
- [9] Michalczyk A. - *Model systemu informatycznego wspierającego zarządzanie wiedzą i doświadczeniem w firmach windykacyjnych*. Praca doktorska, University of Warsaw, Faculty of Management, Warszawa 2007.
- [10] Mulawka J. - *Systemy ekspertowe*. WNT, Warszawa 1997.
- [11] Rutkowski L. - *Metody i techniki sztucznej inteligencji*. PWN, Warszawa 2006.
- [12] Sun Z., Finnie G. - *Experience Management in Knowledge Management* [in] *Proceedings of the Knowledge-Based Intelligent Information and Engineering Systems. 9th International Conference (KES2005)* (ed. R. Khosla, R.J. Howlett, L.C. Jain). Melbourne 2005, pp. 979-986.
- [13] Watson I., Marir F. - *Case-Based Reasoning: A Review* [in] *The Knowledge Engineering Review*, Vol. 9, No. 4, 1994, pp. 327-354.
- [14] Waściński T., Michalczyk A. - *Koncepcja systemu IT wspomagającego zarządzanie wiedzą i doświadczeniem* [in] *Zarządzanie wiedzą i technologiami informatycznymi* (ed. C. Orłowski, Z. Kowalczyk, E. Szczerbicki). PWNT, Gdańsk 2008, pp. 167-174.