Fundamenta Informaticae 59 (2004) 107–118
IOS Press

A View on Rough Set Concept Approximations

Jan Bazan[*]
Institute of Mathematics,University of Rzeszów
Rejtana 16A, 35-959 Rzeszów, Poland
bazan@mimuw.edu.pl

Nguyen Hung Son and Marcin Szczuka
Institute of Mathematics, Warsaw University
Banacha 2, 02-097 Warsaw, Poland
{*son,szczuka*}*@mimuw.edu.pl*

Abstract. The concept of approximation is one of the most fundamental in rough set theory. In this work we examine this basic notion as well as its extensions and modifications. The goal is to construct a parameterized approximation mechanism making it possible to develop multi–stage multi–level concept hierarchies that are capable of maintaining acceptable level of imprecision from input to output.

1. Introduction

The notion of concept approximation is a focal point of many approaches to data analysis based on rough set theory. The original concept of indiscernibility and approximation (Pawlak, 1982), as presented in [12], is meant to provide a way of dealing with inconsistency and incompleteness in data. Elegantly and simply devised, the rough set approximations proved to be a useful tool in supporting data–related tasks such as classification, decision making, and description.

In the original rough set setting we are looking on data using the "glasses" which are determined by the choice of attributes (measurements) we are provided. The inconsistency, vagueness and imprecision are intrinsic to the information system we are given. The shape of approximation (upper/lower) depends

[*]Address for correspondence: Institute of Mathematics, Warsaw University, Banacha 2, 02-097 Warsaw, Poland

only on the choice of attributes. That may be regarded as both advantage, since it provides simplicity and clarity and disadvantage, since there may not be enough flexibility for some applications.

It should be, however, noticed that the canonical approach to concept approximation is rarely used to the full extent. In the majority of rough set applications, the approximations based on the original set of attributes (objects' features) are used only at some initial stage of inductive learning. However, in most cases the final system is based on extended representation. Majority of the existing solutions (see [2, 7, 11]) make use of decision rules derived from data and accompanied with a *recipé* for their usage. We show that approximations of concepts retrieved from such a system can be enriched by using some patterns defined by rule based classifiers or other classifiers. Such approach is much more flexible and better suited for multi–stage reasoning schemes (see [17]) that are discussed further in the paper.

The importance of proper approximation choice becomes much more crucial when it comes to construction of compound decision support and classification systems. These systems create higher level concepts using as building blocks not only attribute values but also previously derived, more primitive concepts. The lower level concepts may be created with use of approximate techniques as well. Therefore, higher level concepts incorporate not only the imprecision resulting from the way they are being constructed but also the imprecision inherited from the lower level concepts. The crucial point is to assure the parameterized space of possible approximations. Then, by proper tuning of parameters we may control the proliferation of imprecision. In rough set terms this corresponds to setting proper constraints for the size and shape of boundary region.

To illustrate the problems that require multi–layered approximation schemes and compound concept approximation, let us bring two examples related to RoboCup [1] and WITAS project [3, 10, 17].

The RoboCup [1] is an initiative aimed at fostering research in the field of cooperative autonomous robotics by construction of programs for control of robots which play football. If we consider the algorithms controlling behavior of the player (robot) we have to take into account several concepts which should be understandable to it. The player perceives its current stage by measuring some attributes as position, position of other robots, position of the ball and behavior of other players. To take decision the player should not only take into account simple observations, but also the concepts of higher level such as situation on the field. The decision rule for such a robot could be:

If *all players of my team are forwarding* ***and*** *all players of the opposite team I can see are backing up* ***then*** *I should go forward* ***even if*** *I have no information about current position of the ball.*

Notice that to use such rule the robot has to establish first concepts such us *all players*, *go forward* etc., and then use them in higher level decision making. Naturally, such more primitive concepts may be approximately established.

Another example comes from the WITAS project (see [3, 10, 17]). The aim of this project is to construct an unmanned aerial vehicle (autonomous helicopter) capable of recognizing the road situation underneath and take appropriate action. The aircraft is equipped with several sensors, most importantly - a video camera. From the image analysis module the decision system receives primitive concepts such as color blobs and lines. Then it has to map these measurements to the primitive concepts such as cars, roads, crossings, etc. Next stage involves creation of more compound concepts like *overtake*. From these concepts some more sophisticated notions may be derived. The final rule may look like:

If *two cars are driving at high speed* ***and*** *there is a sharp turn on the road* ***and*** *one car tries to overtake the other* ***and*** *the road is slippery* ***then*** *I should report dangerous situation* ***even if*** *the weather conditions are good.*

Patient	Age	Sex	Cholesterol	Resting ECG	Heart rate	Sick
p_1	53	M	203	hyp	155	yes
p_2	60	M	185	hyp	155	yes
p_3	40	M	199	norm	178	no
...	...	...	...	...	...	...

Table 1. Example of information table from heart–disease domain

The paper presents the topics sketched above in the following way. First, the concept of rough set approximation is introduced and discussed. After that we present and discuss the approximations induced by rule sets and concept assessment schemes accompanying them. Finally, we present how the proposed approach deals with the issue of construction of compound concept hierarchies while preserving approximation quality. We also make connections with the concept of parameterized approximation space [17] and ideas connected to the approach known as *rough mereology* [14, 15]. This paper is an extension of contribution [4] that was presented at the Conference on Rough Sets, Fuzzy Sets, Data Mining and Granular Computing in Chongqing, China.

2. Rough set preliminaries

2.1. Information systems

An *information system* [12] is a pair $\mathbb{S} = (U, A)$, where U is a non–empty, finite set of *objects* and A is a non–empty, finite set, of *attributes*. Each $a \in A$ corresponds to the function $a : U \to V_a$, where V_a is called the *value set* of a. Elements of U could be interpreted as cases, states, patients, observations, etc.

Given an information system $\mathbb{S} = (U, A)$, we associate with any non–empty set of attributes $B \subseteq A$ *the B-information signature* for any object $x \in U$ by $inf_B(x) = \{(a, a(x)) : a \in B\}$. The set $\{inf_A(x) : x \in U\}$ is called the *A–information set* and it is denoted by $INF(\mathbb{S})$.

The above formal definition of information systems is very general and it covers many different "real information systems", e.g., elementary database systems. For simplification, we will use the simplest form of information systems called "information table". Information system can be implemented as a two–dimensional array (matrix) called an *information table*. In an information table, we usually associate its rows with objects (more precisely information vectors of objects), its columns with attributes and its cells with values of attributes on objects (see Table 1).

In supervised learning problems, objects from the training set are pre–classified into several *categories* or *classes*. To deal with such type of data we use *decision systems* of the form $\mathbb{S} = (U, A, dec)$, where $dec \notin A$ is a distinguished attribute called *decision* and elements of attribute set A are called *conditions*. In practice, decision systems contain a description of a finite sample U of objects from larger (possibly infinite) universe $\mathcal{U}$. Conditions are attributes such that their values are known for all objects from $\mathcal{U}$, but decision is a function defined on the objects from the sample U only. In example from Table 1 we can take last column as describing decision. Usually decision attribute is a characteristic function of an unknown concept or several concepts on a sample of objects. The main problem of learning theory is to generalize the decision function (concept description) that is partially defined on the sample U, to the universe $\mathcal{U}$. Without loss of generality one can assume that the domain V_{dec} of the decision dec is equal to $\{1, \ldots, d\}$. The decision dec determines a partition $\{CLASS_1, \ldots, CLASS_d\}$ of the universe U, where $CLASS_k = \{x \in U : dec(x) = k\}$ is called the k–th *decision class of* $\mathbb{S}$ for $1 \leq k \leq d$. By *class distribution* of any set $X \subseteq U$, we mean the vector $ClassDist(X) = \langle n_1, \ldots, n_d \rangle$, where $n_k = card(X \cap CLASS_k)$ is the number of objects from X belonging to the k–th decision class.

2.2. Concept approximation

In many real–life situations, we are not able to give an exact definition of the concept.

For example, frequently we are using adjectives like "good", "nice", etc., but no one can define exactly such concepts like "nice person", "young man". This concept looks to be easy to define by age, e.g., with the rule:

IF $age(X) \leq a$ **THEN** X is young

but it is very unnatural to say that "X is not young because $age(X) = a + \varepsilon$" for small positive ε. Such uncertain situations are caused by either the lack of information about the concept or by the richness of natural language. There are different approaches to deal with uncertain and vague concepts like multi–valued logics, fuzzy set theory, and rough set theory. Using these approaches, concepts are defined by "multi–valued membership function" instead of classical "binary (crisp) membership relations" (set characteristic functions). In particular, what we want to underline in this paper is that rough set approach offers a way for establishing membership functions that are data–grounded, making them significantly different from other kinds of membership functions that need preliminary assumptions.

In rough set approach, assuming a concept X defined over the universe $\mathcal{U}$ of objects ($X \subseteq \mathcal{U}$), the problem is to find a description of the concept X, which can be expressed in a predefined descriptive language, which is a set of formulae that are interpretable as subsets of $\mathcal{U}$. In general, the problem is to find a description of a concept X in a language $\mathcal{L}_2$ (e.g., consisting of boolean formulae defined over subset of attributes) assuming the concept is definable in another language $\mathcal{L}_1$ (e.g., natural language, or defined by a set of attributes).

Usually, the concept X is specified partially, i.e., values of the characteristic function of X is given only on a small subset $U \subseteq \mathcal{U}$ called the training sample. Such information makes it possible to search for patterns in a given language defining in the training sample sets included (or sufficiently included) into a given concept. Observe that the approximations of a concept cannot be defined uniquely for a given sample of objects. The approximations of the whole concept X are obtained by induction from given information on a sample U of objects (containing some positive examples $X \cap U$ and negative examples $\overline{X} \cap U$). Hence, the quality of such approximations should be verified on new testing objects. Thus we propose to search for concept approximations gradually. Parameterized patterns defined by rough membership functions related to classifiers help to discover relevant patterns on the object universe extended by adding new testing objects. In the paper we present illustrative examples of such parameterized patterns. By tuning parameters of such patterns one can obtain patterns relevant for concept approximation of the extended training sample by testing objects from U^* where $U \subseteq U^* \subseteq \mathcal{U}$.

One should also consider uncertainty that may be present while representing objects. They are described by some features (attributes) which causes some objects to become indiscernible with respect to these features. In practice, objects from $\mathcal{U}$ are perceived by means of information vectors being vectors of attribute values (information signature). In this case, the language $\mathcal{L}$ consists of boolean formulae defined over accessible (effectively measurable) attributes.

Due to bounds on expressiveness of language $\mathcal{L}$ in the universe $\mathcal{U}$, we are forced to find some approximate rather than exact description of a given concept. Rough set methodology for approximation of a concept $X \subseteq \mathcal{U}$, assuming X and $\mathcal{U} - -X$ are known only on a sample $U \subseteq \mathcal{U}$, can be based on finding pairs $\mathbb{P} = (\mathbf{L}, \mathbf{U})$ of object sets in $\mathcal{U}$ satisfying the following conditions:

1. $\mathbf{L}, \mathbf{U}, \mathcal{U} \setminus \mathbf{L}, \mathcal{U} \setminus \mathbf{U}$ are subsets of $\mathcal{U}$ expressible in the language $\mathcal{L}$;
2. $\mathbf{L} \cap U \subseteq X \cap U \subseteq \mathbf{U} \cap U$;
3. the set **L** (**U**) is maximal (minimal) in the family of sets definable in $\mathcal{L}$ satisfying (2).

The sets **L** and **U** are called the *lower approximation* and the *upper approximation* of the concept $X \subseteq \mathcal{U}$ (*generated by its sample on* U), respectively. The set $\mathbf{BN} = \mathbf{U} \setminus \mathbf{L}$ is called the *boundary region of approximation* of X. The set X is called *rough* with respect to its approximations $(\mathbf{L}, \mathbf{U})$ if $\mathbf{L} \neq \mathbf{U}$, otherwise X is called *crisp* in $\mathcal{U}$. the pair $(\mathbf{L}, \mathbf{U})$ is also called the "*rough set*" (for the concept X). In practical applications the last constraint in the above definition can be hard to satisfy. Hence, by using some heuristics we construct sub–optimal instead of maximal or minimal sets. Also, since at the moment of approximation construction we only know U it may be necessary to change approximation after we gain more information about $\mathcal{U}$ (new objects arrive). The rough approximation of concept can be also defined by means of rough membership function.

Definition 2.1. A function $f : \mathcal{U} \to [0, 1]$ is called *a rough membership function* of the concept $X \subseteq \mathcal{U}$ approximated by $(\mathbf{L}, \mathbf{U})$ (assuming X and $\mathcal{U} - X$ are known only on a sample $U \subseteq \mathcal{U}$) if and only if $\mathbf{L} = \mathbf{L_f} = \{x \in \mathcal{U} : f(x) = 1\}$ and $\mathbf{U} = \mathbf{U}_f = \{x \in \mathcal{U} : f(x) > 0\}$.

Let us note some general features of rough membership functions from Definition 2.1:

1. If we are in possession of an algorithm which makes it possible to construct a function $f(x)$ for the set of attributes A on the basis of $inf_A(x)$ then any function $f : \mathcal{U} \to [0, 1]$ established in this way is a rough membership function on $\mathcal{U}$.
2. Any rough membership function $f : \mathcal{U} \to [0, 1]$ defines a rough approximation $\mathbb{P}_f = (\mathbf{L}_f, \mathbf{U}_f)$, where
$$\mathbf{L}_f = \{x \in \mathcal{U} : f(x) = 1\}; \qquad \mathbf{U}_f = \{x \in \mathcal{U} : f(x) > 0\}$$

Example 2.1. Let us consider the function $f : \mathbb{R}^+ \to [0, 1]$ defined as follows:

$$f(x) = \begin{cases} 1 & \text{if } x < 30 \\ 0.5 & \text{if } x \in [30, 50] \\ 0 & \text{if } x > 50 \end{cases}$$

This function can be treated as rough membership function of the notion: "*the young man*". The lower and upper approximation together with and boundary region are as follows:

$$\mathbf{L}_f = [0, 30); \qquad \mathbf{U}_f = [0, 50] \qquad \mathbf{BN}_f = [30, 50]$$

From mathematical point of view, the previous definition does not make a difference between rough membership functions and fuzzy membership functions. The difference is based on the way they are established. The fuzzy membership functions of sets (also set operations like sum, product or negation) are usually determined by expert, while rough membership functions are derived from experimental data.

Note that the proposed concept approximations are not defined uniquely from a given sample $X \cap U$ of objects. They are obtained by inducing the concept $X \subseteq \mathcal{U}$ approximations from given information on $X \cap U$. Hence, the quality of such approximations should be verified on new objects. This is the reason behind our proposal of searching for concept approximations gradually. The parametrization delivered by rough membership functions related to classifiers makes it possible to discover relevant patterns in the object universe extended by adding new (testing) objects. In the following sections we present illustrative examples of such parameterized patterns. By tuning parameters of such patterns one can obtain patterns relevant for concept approximation of the extended training sample by testing objects.

3. Attribute–based approximations

In this section we discuss a basic approach to determine the rough approximations to a concept. It was presented in [12] and will be called herein "standard."

Rough set approximations [12, 13] are fundamental and widely used in many reasoning methods under uncertainty (caused, e.g., by the lack of some attributes). For a given information system $\mathbb{S} = (U, A)$ and an attribute set $B \subseteq A$, one can define a *B-indiscernibility relation* $IND(B)$ assuming

$$IND(B) = \{(x, y) \in U \times U : inf_B(x) = inf_B(y)\}\,.$$

Its equivalence classes are defined by
$[x]_{IND(B)} = \{u \in U : (x, u) \in IND\,(B)\}$ for any object $x \in U$. The problem is to define a concept $X \subseteq U$, assuming that only some attributes from $B \subseteq A$ are given. This problem is often specified by a decision system $\mathbb{S}_1 = (U, B, dec_X)$, where $dec_X(u) = 1$ for $u \in X$, and $dec_X(u) = 0$ for $u \notin X$. Attributes from B determine the rough membership function $\mu_X^B : U \to [0, 1]$ for the concept X by $\mu_X^B(x) = card(X \cap [x]_{IND(B)})/card([x]_{IND(B)})$.

$$\mu_X^B(x) = \frac{card(X \cap [x]_{IND(B)})}{card([x]_{IND(B)})}$$

This function, according to the rough membership function definition, yields rough approximations of the concept X by using indiscernibility classes:

$$\mathbf{L}_B(X) = \mathbf{L}_{\mu_X^B} = \left\{x \in U : \mu_X^B(x) = 1\right\} = \left\{x \in U : [x]_{IND(B)} \subseteq X\right\}$$

$$\mathbf{U}_B(X) = \mathbf{U}_{\mu_X^B} = \left\{x \in U : \mu_X^B(x) > 0\right\} = \left\{x \in U : [x]_{IND(B)} \cap X \neq \varnothing\right\}$$

called the *B–lower* and the *B–upper approximation* of X in $\mathbb{S}$, respectively. The set $\mathbf{BN}_B(X) = \mathbf{U}_B(X) \setminus \mathbf{L}_B(X)$ is called *B–boundary region* of the concept X.

Please note that in such a definition of approximation we adopt the *closed world assumption*, i.e., the concept approximation problem is related to objects from the information system $\mathbb{S}$ only. In inductive learning, it is necessary to extend the rough set based approximations for objects outside U. Unfortunately, the lack of generalization in the process of attribute–based approximations implies that there may exist objects $x \in \mathcal{U} \setminus U$ satisfying $[x]_{IND(B)} \cap U = \varnothing$. Hence, for such objects we are unable to make any decision. In the following sections we discuss some extensions of approximations on supersets of U which are less sensitive to the problems mentioned above.

4. Case–based approximations

In case–based reasoning methods, like k–NN (k–Nearest–Neighbor, see [8]) method, it is necessary to define a distance function between objects $\delta : \mathcal{U} \times \mathcal{U} \to \mathbb{R}^+$. The problem of searching for a relevant distance function for the given data set is not trivial, but at this point, let us assume that such function has been already defined.

In k–NN classification method, the decision for a new object $x \in \mathcal{U} \setminus U$ is made on the basis of the set $NN(x;k) := \{x_1, \ldots, x_k\} \subseteq U$ with k objects from U which are nearest to x with respect to the distance function δ. So, for any object $x_i \in NN(x;k)$ and for any object $u \in U \setminus NN(x;k)$ we have $\delta(x, x_i) \leq \delta(x, u)$. Usually k is a parameter which can be determined by expert or constructed from experimental data. The k-NN classifiers frequently use voting algorithm for establishing final decision, i.e., the decision value for new object x can be predicted by: $dec(x) = Voting(\langle n_1, \ldots, n_d \rangle)$ where $\langle n_1, \ldots, n_d \rangle = ClassDist(NN(x;k))$ is the class distribution of the set $NN(x,k)$ satisfying $n_1 + \ldots + n_d = k$. The voting function can return the most frequent decision value occurring in $NN(x,k)$, i.e., $dec(x) = i$ if and only if n_i is the largest value among $\langle n_1, \ldots, n_d \rangle$. In case of imbalanced data, the vector $\langle n_1, \ldots, n_d \rangle$ can first be scaled with respect to global class distribution, and after that the voting algorithm can be employed.

We are now going to present the rough approximation based on the sets $NN(x;k)$. Let us define a family of functions by

$$\mu^{t_1,t_2}_{CLASS_i}(x) = \begin{cases} 1 & \text{if } n_i \geq t_2 \\ \frac{n_i - t_1}{t2 - t1} & \text{if } n_i \in (t_1, t_2) \\ 0 & \text{if } n_i \leq t_1 \end{cases}$$

where $t_1 < t_2 < k$, n_i is the i–th coordinate in the class distribution of $NN(x;k)$. Any such function defines patterns described by means of the following formulae: $\mu^{t_1,t_2}_{CLASS_i}(x) \circ c$, where $\circ \in \{=, \geq, \leq, <, >\}$ and $c \in \{0, 1, \frac{n_i - t_1}{t_2 - t_1}\}$. One can tune parameters of such formulae to obtain new relevant patterns for the concept approximation on the considered extension of the universe by testing objects.

This rough membership function defines a rough approximation of i^{th} decision class. For an object $x \in \mathcal{U}$, we have:

$$x \in \mathbf{L}_{kNN}(CLASS_i) \Leftrightarrow n_i \geq t_1 \text{ and } x \in \mathbf{U}_{kNN}(CLASS_i) \Leftrightarrow n_i \geq t_2$$

As we mentioned above, the disadvantage of k–NN methods is based on the assumption that the distance function is already defined for all pairs of objects, which is not the case for many complex data sets.

In the next section we present the alternative way to define the rough approximation.

5. Rule–based approximations

Let $\mathbb{S} = (U, A, dec)$ be a decision system. Any implication of the form

$$(a_{i_1} = v_1) \wedge \ldots \wedge (a_{i_m} = v_m) \Rightarrow (dec = k) \quad (1)$$

where $a_{i_j} \in A$ and $v_j \in V_{a_{i_j}}$, is called a *decision rule* for the k–th decision class. Any decision rule $\mathbf{r}$ of the above form can be characterized by the following parameters:

- $length(\mathbf{r})$ = the number of descriptors in the premise of $\mathbf{r}$;
- $[\mathbf{r}]$ = *carrier of* $\mathbf{r}$, i.e., the set of objects satisfying the premise of $\mathbf{r}$;
- $support(\mathbf{r})$ = number of objects satisfying the premise of $\mathbf{r}$;
- $confidence(\mathbf{r})$ = the measure of truth of the decision rule = $\frac{card([\mathbf{r}] \cap CLASS_k)}{card([\mathbf{r}])}$.

  The decision rule $\mathbf{r}$ is called *consistent* with $\mathbb{A}$ if $confidence(\mathbf{r}) = 1$.

In data mining, we are interested in searching for *short, strong* decision rules with *high confidence*. The linguistic features like "short", "strong" or "high confidence" of decision rules can be formulated by means of their length, support or confidence. Many decision rule generation methods have been developed using rough set theory (see e.g., [5, 7, 16, 2]).

One of the most interesting approaches is related to *minimal consistent decision rules*. Given a decision table $\mathbb{S} = (U, A \cup \{dec\})$, the decision rule: $\mathbf{r} =_{def} (a_{i_1} = v_1) \wedge \ldots \wedge (a_{i_m} = v_m) \Rightarrow (dec = k)$ is called minimal consistent decision rule if it is consistent with $\mathbb{S}$ and any decision rule $\mathbf{r}'$ created from $\mathbf{r}$ by removing one of descriptors from left hand side of $\mathbf{r}$ is not consistent with $\mathbb{S}$. The set of all minimal consistent decision rules for a given decision table $\mathbb{S}$, denoted by $MinConsRules(\mathbb{S})$, can be found by computing *object oriented reducts* (or local reducts) as in [5, 9, 19].

The elements of $MinConsRules(\mathbb{S})$ can be treated as interesting, valuable and useful patterns in data and used as a knowledge base in classification systems. Unfortunately, the number of such patterns can be exponential with respect to the size of a given decision table [5, 9, 16, 19]. In practice, we must apply some heuristics to generate a subset of decision rules. We mention some of them:

1. **Rule filtering:** Instead of $MinConsRules(\mathbb{S})$, we can use the set of short, strong, and highly accurate decision rules defined by:

   $$MinRules(\mathbb{S}, \lambda_{\max}, \sigma_{\min}, \alpha_{\min}) = \{\mathbf{r} : \mathbf{r} \text{ is minimal } \wedge length(\mathbf{r}) \leq \lambda_{\max}$$

   $$\wedge support(\mathbf{r}) \geq \sigma_{\min} \wedge confidence(\mathbf{r}) \geq \alpha_{\min}\}$$

   All heuristics for object oriented reducts can be modified to extracting decision rules from $MinRules(\mathbb{S}, \lambda_{\max}, \sigma_{\min}, \alpha_{\min})$.

2. **Object covering:** The main idea is based on searching from $MinConsRules(\mathbb{S})$ minimal set of rules that cover almost all objects from U. There exist many such covering algorithms, see [7, 19] for more details;

The rule based classification methods work in three phases:

1. Learning phase: generates a set of decision rules *RULES*$(\mathbb{S})$ (satisfying some predefined conditions) from a given decision system $\mathbb{S}$.

2. Rule selection phase: selects from *RULES*$(\mathbb{S})$ the subset of rules such that they are supported by x, where $x \in \mathcal{U}$ is a testing object. We denote this set by $MatchRules(\mathbb{S}, x)$.

3. Decision making phase: makes a decision for x using some voting algorithm for decision rules from $MatchRules(\mathbb{S}, x)$

A rule based classifier works as follows. Suppose we would like to decide if a given object $x \in \mathcal{U}$ belongs to the i–th decision class. Let $MatchRules(\mathbb{S}, x) = \mathbf{R}_{yes} \cup \mathbf{R}_{no}$, where $\mathbf{R}_{yes}$ is the set of all decision rules for i-th class matched by x and $\mathbf{R}_{no}$ is the set of decision rules for other classes matched by x. We assign two real values w_{yes}, w_{no} defined by

$$w_{yes} = \sum_{\mathbf{r} \in \mathbf{R}_{yes}} strength(\mathbf{r}) \qquad w_{no} = \sum_{\mathbf{r} \in \mathbf{R}_{no}} strength(\mathbf{r})$$

where w_{yes}, w_{no} are called *for* and *against* weights of the object x, and $strength(\mathbf{r})$ is a normalized function depending on $length(\mathbf{r})$, $support(\mathbf{r})$, $confidence(\mathbf{r})$ and some global information about the decision system $\mathbb{S}$ like decision system size, global class distribution, etc. (see [5]). Using some relationships between w_{yes}, w_{no} the classifier predicts the decision.

Note, that in such an approach any classifier can be identified with a membership function. We can define rule–based classifiers by a parameterized function of the following form:

IF $\max(w_{yes}, w_{no}) < \omega$ **THEN** $\mu_{CLASS_k}(x) = 0$

$$\textbf{ELSE} \quad \mu_{CLASS_k}(x) = \begin{cases} 1 & \text{if } w_{yes} - w_{no} \geq \theta \\ \frac{\theta + (w_{yes} - w_{no})}{2\theta} & \text{if } |w_{yes} - w_{no}| < \theta \\ 0 & \text{if } w_{yes} - w_{no} \leq -\theta \end{cases}$$

where ω and θ are parameters that allow to search for new relevant patterns (pieces of concept description) for the concept approximation on the extension of the initial training sample by testing objects. Proper tuning of ω and θ gives us ability to retain the desired level of confidence without necessity of $\mu_{CLASS_k}(x)$ reconstruction, while new, previously unseen objects appear.

6. Approximations of compound objects

As mentioned before, here we are not only concerned with approximation of concepts that are described with simple attributes but also with higher level concepts established from already existing ones. The idea is to use approximations in the way which gives us the ability to control the level of approximation quality (imprecision level).

To simplify notation let us assume that we have two concepts C_1 and C_2 that are given by means of rule–based approximations derived from decision systems $\mathbb{S}_{C_1}$ and $\mathbb{S}_{C_2}$, where $\mathbb{S}_{C_1} = (U, A_{C_1}, dec_{C_1})$ and $\mathbb{S}_{C_2} = (U, A_{C_2}, dec_{C_2})$.

Hence, we are given two sets of patterns for approximation of concepts C_1 and C_2 (see Section 5). These patterns are defined by the left hand sides of minimal decision rules formulae related to the classifiers used for construction of approximations for C_1 and C_2 . They can be obtained by tuning parameters

$\{w_{yes}^{C_1}, w_{no}^{C_1}, \omega^{C_1}, \theta^{C_1}\}$ and $\{w_{yes}^{C_2}, w_{no}^{C_2}, \omega^{C_2}, \theta^{C_2}\}$ discussed previously. We want to establish a relevant set of patterns and parameters for the target concept C, i.e. $\{w_{yes}^{C}, w_{no}^{C}, \omega^{C}, \theta^{C}\}$.

Henceforth we are given two rough membership functions $\mu_{C_1}(x)$, $\mu_{C_2}(x)$. These functions are determined with use of parameter sets $\{w_{yes}^{C_1}, w_{no}^{C_1}, \omega^{C_1}, \theta^{C_1}\}$ and $\{w_{yes}^{C_2}, w_{no}^{C_2}, \omega^{C_2}, \theta^{C_2}\}$, respectively.

Similar set of parameters $\{w_{yes}^{C}, w_{no}^{C}, \omega^{C}, \theta^{C}\}$ for the target concept C, which we want to describe with use of rough membership function $\mu_C(.)$. As previously, the parameters ω, θ are user-configurable and serve as controls for the boundary region. But, we still need to derive $\{w_{yes}^{C}, w_{no}^{C}\}$ from the data.

The issue is to define a decision system from which we can derive rules determining approximations of C. Let us recall that both simpler concepts C_1, C_2 and the target concept C are defined over the same universe $\mathcal{U}$ and are specified on a sample $U \subseteq \mathcal{U}$. To complete the construction of $\mathbb{S}_C = (U, A_C, dec_C)$ we need to specify $A_C \cup \{dec_C\}$. The decision attribute is known for an arbitrary object $x \in U$ and conditional attributes in (the simplest case of) our proposal are either rough memberships for simpler concepts ($A = \{\mu_{C_1}(x), \mu_{C_2}(x)\}$) or weights for simpler concepts ($A = \{w_{yes}^{C_1}, w_{no}^{C_1}, w_{yes}^{C_2}, w_{no}^{C_2}\}$). In the former case we concentrate on the degree of inclusion while in the later case we take into account the relationships with positive and negative information generated during object classification.

By performing learning of rules from $\mathbb{S}_C$ we create rule–based approximations of the concept C. It is important to mention that such rules describing C can use attributes that are in fact outputs of classifiers themselves. Therefore, in order to have more readable and intuitively understandable description as well as more control over quality of approximation (especially for new cases) it pays off to stratify and interpret attribute domains for attributes in A_C. Relevant patterns used in rules for such approach can then match linguistic statements such as *"the likeliness of the occurrence of C_1 is low"*. Subsets of attribute values are then identified with notions such us *"certain"*, *"low"*, *"high"* etc. It is quite natural, to introduce linearly ordered subsets of attribute ranges, e.g., $\{negative, low, medium, high, positive\}$. That yields fuzzy–like layout of attribute values. One may (and in some cases should) consider also the case when these subsets overlap. Then, there may be more linguistic valued attached to attribute values since variables like *low* or *medium* appear.

Stratification of attribute values and introduction of linguistic variable attached to the strata provides a way for representing knowledge in more human-readable format since for new object $x^* \in \mathcal{U} \setminus U$ to be classified we may use rules like:

If *compliance of x^* with C_1 is <u>high</u> or <u>medium</u>* ***and*** *compliance of x^* with C_2 is <u>high</u>* ***then*** $x^* \in C$.

Another advantage of imposing the division of attribute value sets lays in extended control over flexibility and validity of system constructed in this way. We gain the ability of making system more stable and inductively correct and we control the general layout of boundary regions that contribute to construction of the target concept. The process of setting the intervals for attribute values may be performed by hand or with use of automated methods for interval construction, e.g., clustering, template analysis, and discretization. For some discussion of this approach, related to *rough neurocomputing* and *computing with words* see [17].

7. Conclusions and further directions

In the paper we have presented a collection of basic ideas that redefine the view on the approximation of concepts in rough set framework. There is still much more to be done. Only some of the techniques mentioned in the paper are already implemented and can be tested (see e.g., [2]).

Still, there are much more issues that accompany the task of proper approximation construction. One of them is related to an incremental construction (incremental learning) of approximations. It is a big challenge to devise the method that will be capable of learning an approximation from the sample U and then, given enriched (finite) sample $U^* \supset U$, extend the approximation with few simple steps, without fundamental reconstruction.

Another topic is the investigation of partially defined rough membership functions. In this paper the set of objects, which do not belong to the rough membership function's domain, is treated as a part of the boundary region (see sect. 5). This solution stems from the fact that rough set methods are based on intuitive 3–valued logics. We are convinced that it can be improved by introducing intuitive 4-value logics [6, 18].

Acknowledgements:

The research has been supported by the grant 3T11C00226 from Ministry of Scientific Research and Information Technology of the Republic of Poland.

References

[1] The RoboCup Homepage, `http://www.robocup.org`.

[2] The RSES Homepage, `http://logic.mimuw.edu.pl/∼rses`.

[3] The WITAS Homepage, `http://www.ida.liu.se/ext/witas/`.

[4] Bazan, J., Son, N. H., Skowron, A., Szczuka, M. S.: A View on Rough Set Concept Approximations, *Proceedings of RSFDGrC'2003* (G. Wang, Q. Liu, Y. Yao, A. Skowron, Eds.), LNAI 2639, Springer-Verlag, Berlin, 2003.

[5] Bazan, J. G.: A comparison of dynamic non-dynamic rough set methods for extracting laws from decision tables, in: *Rough Sets in Knowledge Discovery* (L. Polkowski, A. Skowron, Eds.), Physica-Verlag, Heidelberg, 1998, 321–365.

[6] Belnap, N. D.: A useful four-valued logic, in: *Modern uses of multiple-valued logic* (J. M. Dunn, G. Epstein, D. Reidel, Eds.), Reidel Publishing Company, Boston, 1977, 8–37.

[7] Grzymała-Busse, J.: A new version of the rule induction system LERS, *Fundamenta Informaticae*, **31**(1), 1997, 27–39.

[8] Hastie, T., Tibshirani, R., Friedman, J.: *The elements of statistical learning*, Springer Series in Statistics, Springer-Verlag, New York, 2001.

[9] Komorowski, J., Pawlak, Z., Polkowski, L., Skowron, A.: Rough sets: A tutorial, in: *Rough - fuzzy hybridization: A new trend in decision making* (S. K. Pal, A. Skowron, Eds.), Springer-Verlag, Singapore, 1999, 3–98.

[10] Nguyen, H. S., Skowron, A., Szczuka, M.: Situation identification by unmanned aerial vehicle, *Rough Sets and Current Trends in Computing (Proc. RSCTC'2000, Banff, Canada)* (W. Ziarko, Y. Yao, Eds.), LNAI 2005, Springer-Verlag, Banff, Canada, 2001.

[11] Ohrn, A., Komorowski, J., Skowron, A., Synak, P.: The ROSETTA software system, in: *Rough Sets in Knowledge Discovery* (L. Polkowski, A. Skowron, Eds.), Physica-Verlag, Heidelberg, 1998, 572–576.

[12] Pawlak, Z.: *Rough sets: Theoretical aspects of reasoning about data*, Kluwer Academic Publishers, Dordrecht, 1991.

[13] Pawlak, Z., Skowron, A.: Rough membership functions, in: *Advances in the Dempster–Shafer theory of evidence* (R. Yager, M. Fedrizzi, J. Kacprzyk, Eds.), Wiley, New York, 1994, 251–271.

[14] Polkowski, L., Skowron, A.: Rough mereology: A new paradigm for approximate reasoning, *Journal of Approximate Reasoning*, **15**(4), 1996, 333–365.

[15] Polkowski, L., Skowron, A.: Towards an adaptive calculus of granules, in: *Computing with Words in Information/Intelligent Systems* (L. A. Zadeh, J. Kacprzyk, Eds.), Physica-Verlag, Heidelberg, 1999, 201–228.

[16] Rauszer, C., Skowron, A.: The discernibility matrices and functions in information systems, in: *Intelligent decision support – Handbook of applications and advances of the rough set theory* (R. Słowiński, Ed.), Kluwer Academic Publishers, Dordrecht, 1992, 311–362.

[17] Skowron, A., Szczuka, M.: Approximate reasoning schemes: Classifiers for Computing with Words, *Proceedings of the conference Soft Methods in Probability, Statistics and Data Analysis (SMPS2002)* (P. Grzegorzewski, O. Hryniewicz, M. Gil, Eds.), Advances in Soft Computing, Physica-Verlag, Heidelberg, 2002.

[18] Vitoria, A., Małuszyński, J.: A logic programming framework for rough sets, *Proceedings of RSCTC'2002* (J. Alpigini, J. F. Peters, A. Skowron, Eds.), LNAI 2475, Springer-Verlag, Berlin, 2002.

[19] Wróblewski, J.: Covering with reducts - a fast algorithm for rule generation, *Proceedings of RSCTC'1998* (L. Polkowski, A. Skowron, Eds.), LNAI 1424, Springer-Verlag, Berlin, 1998.