Vague Concepts: A Rough Set Approach

Andrzej Skowron
Institute of Mathematics, Warsaw University
Banacha 2, 02-097 Warsaw, Poland
skowron@mimuw.edu.pl

Abstract The approximation space definition has evolved in rough set theory over the last 15 years. The aim was to build a unified framework for concept approximations. We present an overview of this evolution together with some operations on approximation spaces that are used in searching for relevant approximation spaces. Among such operations are inductive extensions and granulations of approximation spaces. We emphasize important consequences of the paper for research on approximation of vague concepts and reasoning about them in the framework of adaptive learning. This requires developing new approach to vague concepts going beyond the traditional rough or fuzzy approaches.

Keywords: Vague Concept Approximation, Reasoning about Vague Concepts, Approximation Spaces, Rough Sets, Inductive Extensions, Granulation of Approximation Spaces, Adaptive Learning.

1 Introduction

We present an overview of the evolution of approximation spaces in rough set theory [9] emphasizing several aspects related to the containment relation, information granulation, and some operations on approximation spaces. Our discussion begins from the basic case with the crisp containment of sets [9]. We also include examples of containment relations used in the definition of more advanced approximation spaces [17, 18, 20]. Information granulation is another important aspect that has influenced the definition of approximation spaces. To illustrate this we start from the approximation space definition based on elementary information granules defined by indiscernibility neighborhoods of objects [9]. Next, we recall [17] the approximation spaces consisting of neighborhoods of objects (elementary information granules) and inclusion functions. Finally, we outline the process of granulation of approximation spaces based on granulation of relational structures representing the structure of objects [11, 18, 20].

Among operations on approximation spaces, used in searching for relevant approximation spaces for concept approximations, are inductive extensions. They are important in dealing with imperfect knowledge and vague concepts. There is a long debate in philosophy on vague concepts (see, e.g., [7]). Nowadays, computer scientists are also interested in vague (imprecise) concepts. Lotfi Zadeh [28] introduced a very successful approach to vagueness. In this approach sets are defined by partial membership, in contrast to crisp membership used in the classical definition of a set. Rough set theory [9] expresses vagueness, not by means of membership, but by employing the boundary region of a set. If the boundary region of a set is empty it means that a particular set is crisp, otherwise the set is rough (inexact). The non-empty boundary region of the set means that our knowledge about the set is not sufficient to define the set precisely. Discussion on vagueness in the context of fuzzy sets and rough sets can be found in [16]. In this paper some consequences on understanding of vague concepts caused by inductive extensions of approximation spaces and adaptive concept learning are also outlined.

2 Approximation spaces

In [9] any approximation space is defined as a pair (U, R), where U is a universe of objects and $R \subseteq U \times U$ is an indiscernibility relation defined by an attribute set. The definition of R-approximations of any set $X \subseteq U$, where R is an equivalence relation, is based on the exact (crisp) containment of sets, i.e.,

1. if $[x]_R \subseteq X$, then the object $x \in U$ belongs (with certainty) to the set $X \subseteq U$ (i.e., x belongs to the R-lower approximation of X);
2. if $[x]_R \subseteq U - X$, then the object $x \in U$ belongs (with certainty) to the complement of the set $X \subseteq U$ (i.e., x belongs to the R-lower approximation of $U - X$);
3. if $[x]_R \cap X \neq \emptyset$ and $[x]_R \cap (U - X) \neq \emptyset$, then x belongs (with certainty) to the R-boundary region of X;

where $[x]_R$ denotes the equivalence class of R defined by x.

Approximation spaces were generalized by introducing indiscernibility relations based on similarity (tolerance) relations (see, e.g., [17, 21, 2, 24]).

Approximations of concepts (identified with subsets of U) were initially defined using the (exact) containment $\nu : P(U) \times P(U) \longrightarrow \{0, 1\}$, where

$$\nu(X, Y) = \begin{cases} 1 & \text{if} \quad X \subseteq Y \\ 0 & \quad \text{otherwise.} \end{cases} \tag{1}$$

for any $X, Y \subseteq U$.

The first attempt to relax this constraint in rough set theory was presented in [10, 27]. This was generalized in approximation spaces introduced in [17] where the containment can be of a much more general form. Moreover, in this paper the generalization of object neighborhood (object granule [5]) was introduced.

Let us recall the definition of an approximation space from [17]. For simplicity of reasoning we omit parameters labelling components of approximation spaces that are tuned in searching for relevant approximation spaces.

Definition 1 *An approximation space is a tuple $AS = (U, I, \nu)$, where*

- *U is a non-empty finite set of objects,*
- *$I : U \to P(U)$ is an uncertainty function such that $x \in I(x)$ for any $x \in U$,*
- *$\nu : P(U) \times P(U) \to [0, 1]$ is a rough inclusion function.*

A set $X \subseteq U$ is *definable in AS* if and only if it is a union of some values of the uncertainty function.

The studies on generalization of inclusion functions lead to rough mereology (see, e.g., [14, 15]) where some general constraints are formulated for such functions called rough inclusions. Notice, that it is also possible to consider more soft uncertainty functions defining soft neighborhoods for objects.

A typical example of rough inclusion is the standard rough inclusion function ν_{SRI} defining the degree of inclusion between two subsets of U by

$$\nu_{SRI}(X, Y) = \begin{cases} \frac{card(X \cap Y)}{card(X)} & \text{if} \quad X \neq \emptyset \\ 1 & \text{if} \quad X = \emptyset. \end{cases} \tag{2}$$

Let us note that such an inclusion measure expressed by the confidence coefficient, widely used in data mining [4], has been considered by Łukasiewicz [6] long time ago in studies on assigning the fractional truth values to logical formulas.

The lower and the upper approximations of subsets of U are defined as follows.

Definition 2 *For any approximation space $AS = (U, I, \nu)$, $0 \leq p < q \leq 1$, and any subset $X \subseteq U$ the q-lower and the p-upper approximation of X in AS are defined by*

$$LOW_q(AS, X) = \{x \in U : \nu(I(x), X) \geq q\}, \tag{3}$$

$$UPP_p(AS, X) = \{x \in U : \nu(I(x), X) > p\}, \tag{4}$$

respectively.

From Definition 2 it follows that $LOW_q(AS, X) \subseteq UPP_p(AS, X)$. In general, $LOW_q(AS, X) \subseteq X \subseteq UPP_p(AS, X)$ does not hold. The degree of truth of the following rules is tuned using the parameters p, q:

- if $x \in LOW_q(AS, X)$, then $x \in X$ (i.e., knowing that $x \in LOW_q(AS, X)$ holds there is a high chance that $x \in X$);
- if $x \in U - UPP_p(AS, X)$, then $x \notin X$ (i.e., knowing that $x \in U - UPP_p(AS, X)$ holds there is a high chance that $x \notin X$);
- if $x \in UPP_p(AS, X) - LOW_q(AS, X)$, then x belongs to the boundary region of X (i.e., knowing that $x \in UPP_p(AS, X) - LOW_q(AS, X)$ holds neither $x \in X$ nor $x \in U - X$ can be eliminated).

Approximation spaces can be constructed directly from information systems or from information systems enriched by some similarity relations on attribute value vectors. The above definition generalizes several approaches existing in the literature such as the approximation spaces defined by equivalence or tolerance indiscernibility relations as well as those defined by exact or partial inclusion of indiscernibility classes into concepts [9, 27]. Approximation spaces for information granule systems are investigated in [20].

3 Nonstandard inclusions

In this section we discuss two examples of nonstandard inclusions [20]. The first example is related to function approximation and the second example to extension of inclusion measures on more complex information granules for objects with internal structures. Let us note that in machine learning, pattern recognition, statistical learning [3], and rough set theory [9] one can find many other such examples of inclusion measures, e.g., based on entropy or positive regions that are defined on partitions. More compound information granules lead to other examples of extension of inclusion measures [18].

3.1 Function approximation

One can directly apply the definition of set approximation to relations. For simplicity, but without loss of generality, we consider binary relations only. Let $R \subseteq U \times U$ be a binary relation. One can define approximations of R by an approximation space $AS = (U \times U, I, \nu)$ using a slightly modified Definition 2:

$$LOW_q(AS, R) = \{(x, y) \in U \times U : \nu(I(x, y), X) \geq q\}, \tag{5}$$

$$UPP_p(AS, R) = \{(x, y) \in U \times U : \nu(I(x, y), X) > p\}, \tag{6}$$

for $0 \leq p < q \leq 1$.

The main problem is how to construct relevant approximation spaces, i.e., how to define uncertainty and inclusion functions. One can define, e.g., the uncertainty function by $I(x, y) = I(x) \times I(y)$ (assuming that a one dimensional uncertainty function is given) and the standard inclusion by $\nu = \nu_{SRI}$. Now, let us consider an approximation space $AS = (U, I, \nu)$ and a function $f : Dom \longrightarrow U$, where $Dom \subseteq U$. By $Graph(f)$ we denote the set $\{(x, f(x)) : x \in Dom\}$. One can easily see that if we apply the above definition of relation approximation to f (f is a special case of relation) then the lower approximation is almost always empty. Thus, we have to construct the relevant approximation space $AS^* = (U \times U, I^*, \nu^*)$ in a different way. We assume that the value $I^*(x, y)$ of the uncertainty function, called the neighbourhood (or the window) of (x, y), for $(x, y) \in U \times U$, is defined by

$$I^*(x, y) = I(x) \times I(y). \tag{7}$$

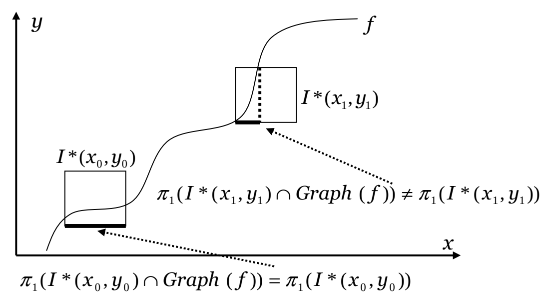

Figure 1: Function approximation

Next, we should decide how to define values of the inclusion function on pairs $(I^*(x, y), Graph(f))$, i.e., how to define the degree r to which the intersection $I^*(x, y) \cap Graph(f)$ is included into $Graph(f)$.

One can consider a ratio

$$r = \frac{card(\{x \in I(x) \cap Dom : f(x) \in I(y)\})}{card(I(x))}, \tag{8}$$

i.e., the ratio of the number of all objects from $I(x) \cap Dom$ (if any) on which f takes a value from $I(y)$ to the number of objects in $I(x)$.

If $r = 1$ then (x, y) defining the window $I^*(x, y)$ is in the lower approximation of $Graph(f)$; if $0 < r \leq 1$ then (x, y) defining the window $I^*(x, y)$ is in the upper approximation of $Graph(f)$.

Using the above intuition, we assume that the inclusion holds to degree one if the domain of $Graph(f)$ restricted to $I(x)$ is equal to $I(x)$ and the values of f on $I(x)$ are in $I(y)$. This can be formally defined by the following condition:

$$\pi_1(I^*(x, y) \cap Graph(f)) = \pi_1(I^*(x, y)), \tag{9}$$

where π_1 denotes the projection on the first coordinate. It is equivalent to

$$\forall x' \in I(x)\ f(x') \in I(y). \tag{10}$$

Thus, the inclusion function ν^* for subsets $X, Y \subseteq U \times U$ is defined by

$$\nu^*(X,Y) = \begin{cases} \frac{card(\pi_1(X \cap Y))}{card(\pi_1(X))} & \text{if} \quad \pi_1(X) \neq \emptyset \\ 1 & \text{if} \quad \pi_1(X) = \emptyset. \end{cases} \tag{11}$$

Hence, the relevant inclusion function in approximation spaces for function approximations is a function that does not measure the degree of inclusion of its arguments but their perceptions, which is represented in the above example by projections of corresponding sets. One can choose another definition, e.g., based on the density of pixels (in case of images) in a window that are matched by the function graph.

Then we have the following proposition, for arbitrary parameters p, q satisfying $0 \leq p < q \leq 1$:

Proposition 1 *Let* $AS^* = (U \times U, I^*, \nu^*)$ *be an approximation space with* I^*, ν^* *defined by* (7), (11), *respectively, and let* $f : Dom \longrightarrow U$ *where* $Dom \subseteq U$. *Then*

1. $(x, y) \in LOW_q(AS^*, Graph(f))$ iff
$$card(\{x' \in I(x) : f(x') \in I(y)\}) \geq q \cdot card(I(x)).$$
2. $(x, y) \in UPP_p(AS^*, Graph(f))$ iff
$$card(\{x' \in I(x) : f(x') \in I(y)\}) > p \cdot card(I(x)).$$

For $q = 1$, $p = 0$ the lower approximation of $G(f)$ consists of all points defining neighborhoods totally covering restrictions of $G(f)$ to these neighborhoods, while to the upper approximation belongs all points defining neighborhoods consisting of at least one point of $G(f)$.

3.2 Relational structure granulation

In this section we discuss an important role that the relational structure granulation [18, 11] plays in searching for relevant patterns in approximate reasoning, e.g., in searching for relevant approximation patterns (see Figure 2). For any object there is defined a neighborhood specified by the value of uncertainty function from an approximation space (see Definition 1). From these neighborhoods some other, more relevant ones (e.g., for the considered concept approximation), should be found. Such neighborhoods can be extracted by searching in a space of neighborhoods generated from values of uncertainty function by applying to them some operations like generalization operations, set theoretical operations (union, intersection), clustering and operations on neighborhoods defined by functions and relations from an underlying relational structure (used to represent an internal object structure).[1] Figure 2 illustrates an exemplary scheme of searching for neighborhoods (patterns, clusters) relevant for concept approximation. In the example f denotes a function with two arguments from the underlying relational structure. Due to the uncertainty, we cannot perceive objects exactly but only by using available neighborhoods defined by the uncertainty function from an approximation space. Hence, instead of the value $f(x, y)$ for a given pair of objects (x, y), one should consider a family of neighborhoods $\mathcal{F} = \{I(f(x', y')) : (x', y') \in I(x) \times I(y)\}$. From this family $\mathcal{F}$, a subfamily $\mathcal{F}'$ of neighborhoods (definable in a given language) can be selected that consists of neighborhoods with some properties relevant for approximation. Next, such subfamily $\mathcal{F}'$ can be generalized to clusters that are relevant for the concept approximation, i.e., clusters included into the approximated concept to a satisfactory degree (see Figure 2). The inclusion degrees can be measured by granulation of a given inclusion function. This requires its extension on the constructed more compound information granules (neighborhoods). We have presented an important construction based on in-

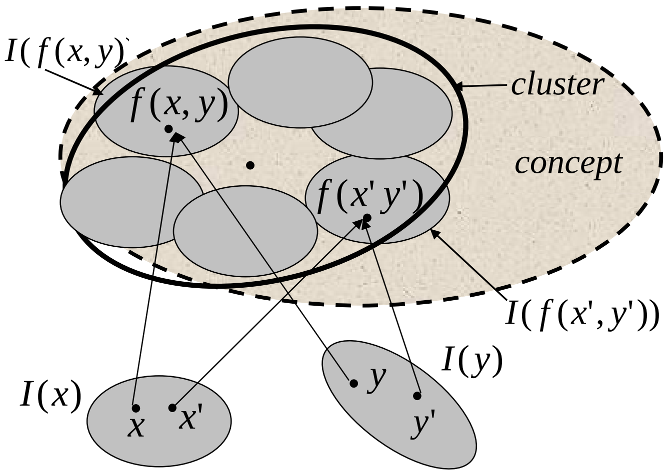

Figure 2: Relational structure granulation

formation granulation used in searching for relevant approximation spaces for concept approximation. In the following section we discuss other important operations on approximation spaces, namely inductive extensions.

4 Concept approximation

In this section we consider the problem of approximation of concepts over a universe U^*, i.e., subsets of U^*. We assume that the concepts are perceived

[1] Relations from such structure may define relations between objects or their parts.

only through some subsets of U^*, called samples. This is a typical situation in machine learning, pattern recognition, or data mining [4]. In this section we explain the rough set approach to induction of concept approximations.

Let $U \subseteq U^*$ be a finite sample and let $C_U = C \cap U$ for any concept $C \subseteq U^*$. Let $AS = (U, I, \nu)$ be an approximation space over the sample U.[2] The problem we consider is how to extend the approximations of C_U defined by AS to approximation of C over U^*. We show that the problem can be described as searching for an extension $AS^* = (U^*, I^*, \nu^*)$ of the approximation space AS, relevant for approximation of C. This requires showing how to induce values of the extended inclusion function to relevant subsets of U^* that are suitable for the approximation of C. Observe (see Definition 2) that for the approximation of C it is enough to induce the necessary values of the inclusion function ν^* without knowing the exact value of $I^*(x) \subseteq U^*$ for $x \in U^*$.

In the following subsections we consider examples for well known classifiers (rule-based classifiers, k-NN classifiers, feed-forward neural networks, and hierarchical classifiers [3]) and we show that their construction is based on the inductive inclusion extension.

4.1 Rule-based classifiers

Let AS be a given approximation space for C_U and let us consider a language L in which the neighborhood $I(x) \subseteq U$ is expressible by a formula $pat(x)$, for any $x \in U$. It means that $I(x) = \|pat(x)\|_U \subseteq U$, where $\|pat(x)\|_U$ denotes the meaning of $pat(x)$ restricted to the sample U. In case of rule based classifiers patterns of the form $pat(x)$ are defined by feature value vectors.

We assume that for any new object $x \in U^* \setminus U$ we can obtain (e.g., as a result of sensor measurement) a pattern $pat(x) \in L$ with semantics $\|pat(x)\|_{U^*} \subseteq U^*$. However, the relationships between information granules over U^*, e.g., $\|pat(x)\|_{U^*}$ and $\|pat(y)\|_{U^*}$, for different $x, y \in U^*$, are known only to a degree estimated by using relationships between the restrictions of these sets to the sample U, i.e., between sets $\|pat(x)\|_{U^*} \cap U$ and $\|pat(y)\|_{U^*} \cap U$.

The set of patterns $\{pat(x) : x \in U\}$ is usually not relevant for approximation of the concept $C \subseteq U^*$. Such patterns can be too specific or not general enough, and can directly be applied only to a very limited number of new objects. However, by using some generalization strategies, one can search in a family of patterns definable from $\{pat(x) : x \in U\}$ in L, for such new patterns that are relevant for approximation of concepts over U^*. Let us consider a subset $PATTERNS(AS, L, C) \subseteq L$ chosen as a set of pattern candidates for relevant approximation of a given concept C. For example, in case of rule based classifier one can search for such candidate patterns among sets definable by subsequences of feature value vectors corresponding to objects from the

[2] For simplicity of reasoning, in this section we use standard definition of approximation spaces (Definition 1).

sample U. The set $PATTERNS(AS, L, C)$ can be selected using some quality measures evaluated on meanings (semantics) of patterns from this set restricted to the sample U (like the numbers of examples from the concept C_U and its complement that support a given pattern). Then, on the basis of properties of sets definable by these patterns over U we induce approximate values of the inclusion function $\nu^*(X, C)$ on subsets of $X \subseteq U^*$ definable by any such pattern and the concept C. Next, we induce the value of ν^* on pairs (X, Y) where $X \subseteq U^*$ is definable by a pattern from $\{pat(x) : x \in U^*\}$ and $Y \subseteq U^*$ is definable by a pattern from $PATTERNS(AS, L, C)$.

Finally, for any object $x \in U^* \setminus U$ we induce the approximation of the degree $\nu^*(\|pat(x)\|_{U^*}, C)$ applying a conflict resolution strategy $Conflict\_res$ (e.g, a voting strategy) to two families of degrees:

$$\{\nu^*(\|pat(x)\|_{U^*}, \|pat\|_{U^*}) : pat \in PATTERNS(AS, L, C)\}, \tag{12}$$

$$\{\nu^*(\|pat\|_{U^*}, C) : pat \in PATTERNS(AS, L, C)\}. \tag{13}$$

Values of the inclusion function for the remaining subsets of U^* can be chosen in any way – they do not have any impact on the approximations of C. Moreover, observe that for the approximation of C we do not need to know the exact values of uncertainty function I^* – it is enough to induce the values of the inclusion function ν^*. The defined extension ν^* of ν to some subsets of U^* makes it possible to define an approximation of the concept C in a new approximation space AS^* by using Definition 2.

Observe, that the value $\nu^*(I^*(x), C)$ of the induced inclusion function for any object $x \in U^* - U$ is based on collected arguments *for* and *against* belonging of x to C. In this way, the approximation of concepts over U^* can be explained as a process of searching for relevant approximation spaces, in particular inducing relevant approximation spaces.

4.2 k-NN classifiers

For simplicity of notation, we consider binary classifiers for a concepts C over U^* and decision tables representing training samples with binary decisions. Let L be a language of patterns given as in the case of rule-based classifiers. We also assume that there is given a distance metric defined between patterns from L and a fixed integer $k \geq 1$. For any object $x \in U^*$ the set $NN(k, x)$ of k nearest to $pat(x)$ patterns (with respect to the given metric function) is extracted. Let $r = card(\{y \in NN(k, x) : \nu(\|y\|_U, C_U) = 1\})$ and $s = card(\{y \in NN(k, x) : \nu(\|y\|_U, U - C_U) = 1\})$. Then the following rule based on majority voting is induced:

$$\begin{aligned} &\text{if } r = s \text{ than } \nu^*(I^*(x), C) = unknown; \\ &\text{if } r > s \text{ than } \nu^*(I^*(x), C) = 1 \text{ else } \nu^*(I^*(x), U^* - C) = 1. \end{aligned} \tag{14}$$

Hence, a given object $x \in U^*$ is classified on the basis of the number of arguments *for* and *against* its belonging to C. Using the induced inclusion function one can induce the lower and the upper approximation of C by

$$LOW(AS^*, C) = \{x \in U^* : \nu^*(I^*(x), C) = 1\};$$
$$UPP(AS^*, C) = \{x \in U^* : \nu^*(I^*(x), C) \in \{1, \text{ unknown}\}\}. \tag{15}$$

The quality of such an approximation can be tested by comparing a testing sample the induced approximations with the real ones, i.e., obtained by extension of a given (training) decision table by adding testing objects. The parameters that can be tuned to obtain the high quality approximations are, in particular, the distance metric, number k, and voting strategy.

4.3 Neural networks

Let us consider the feed-forward neural networks with one hidden layer. Let $A = \{a_1, \ldots, a_m\}$ be a set of real valued attributes defined on the universe U^*. For any object $x \in U^*$ we assume that $pat(x) = (a_1(x), \ldots, a_m(x))$ and $I^*(x) = \{y \in U^* : pat(x) = pat(y)\}$. Now, let us consider patterns defined by vectors $\overline{w}$ from R^m. The semantics of such patterns is defined by $\|\overline{w}\| = f_{\overline{w}} : R^m \longrightarrow R$, where R is the set of reals, $f_{\overline{w}}(x_1, \ldots, x_m) = h(\sum_{i=1}^m w_i x_i)$, and h is a sigmoidal function. The inclusion between the defined kinds of patterns is defined by

$$\nu^*(I^*(x), f_{\overline{w}}) = f_{\overline{w}}(a_1(x), \ldots, a_m(x)), \tag{16}$$

for any object $x \in U^*$.

To estimate for a given classification $\{C_1, \ldots, C_r\}$ the values of inclusion function $\nu^*(I^*(x), C_i)$ for $i = 1, \ldots, r$ and $x \in U^*$ a number l of vectors $\overline{w}_1, \ldots, \overline{w}_l$ (of dimension m) and r vectors $\overline{v}^{(1)}, \ldots, \overline{v}^{(r)}$ (of dimension l) are chosen and the value $\nu^*(I^*(x), C_i)$ is induced using the sum $sgn\{\sum_{j=1}^l v_j^{(i)} f_{\overline{w}_j}(x)\}$ for $i = 1, \ldots, r$. The above sum makes it possible to fuse the arguments *for* and *against* belonging of objects to the decision classes. The parameters to be tuned are, in particular, $\overline{v}^{(1)}, \ldots, \overline{v}^{(r)}$, l, and $\overline{w}_1, \ldots, \overline{w}_l$.

4.4 Hierarchical learning

In case of hierarchical (layered) learning (see, e.g., [22]) searching for arguments *for* and *against* is more compound. The degree to which a given input pattern belongs to the target concept at the highest level in hierarchy (and to its complement) is derived from information on degrees to which such pattern matches patterns at the higher level, degrees to which the matched patterns at this level are matching the patterns at the next higher layer, etc. However, the basic idea of the approximation space extension remains the same.

5 Adaptive Learning as a Basis for Approximate Reasoning about Vague Concepts

We have recognized that for a given concept $C \subseteq U^*$ and any object $x \in U^*$ instead of crisp decision about the relationship of $I^*(x)$ and C we can gather some arguments *for* and *against* it only. Next, it is necessary to induce from such arguments the value $\nu^*(I(x), C)$ using some strategies making it possible to resolve conflicts between those arguments [3, 18]. Usually some general principles are used such as the minimal length principle [3] in searching for algorithms computing an extension $\nu^*(I(x), C)$. However, often the approximated concept is over $U^* - U$ too compound to be induced directly from $\nu(I(x), C)$. This is the reason that the existing learning methods can be not satisfactory for inducing high quality concept approximations in case of complex concepts [25]. There have been several attempts trying to omit this drawback. One of them is the incremental learning used in machine learning and also by rough set community (see, e.g., [26]). In this case an increasing sequence of samples $U_1 \subseteq \ldots \subseteq U_k \subseteq \ldots$ is considered and the task is to induce the extensions $\nu^{(1)}, \ldots, \nu^{(k)}, \ldots$ of inclusion functions. Still we know rather very little about relevant strategies for inducing such extensions. Some other approaches are based on hierarchical (layered) learning [22] or reinforcement learning [23]. However, there are several issues, important for learning that are not within the scope of these approaches. For example, the target concept can gradually change over time and this concept drift is a natural extension for incremental learning systems toward adaptive systems. In adaptive learning it is important not only what we learn but also how we learn, how we measure changes in distributed environment and induce from them adaptive changes of constructed concept approximations. The adaptive learning for autonomous systems became a challenge for machine learning, robotics, complex systems, and multiagent systems. It is becoming also a very attractive research area for the rough set approach. Several results were reported showing that approximate reasoning in distributed environments can be based on rough mereological and granular approaches (see, e.g., [14, 15, 18]). Moreover, investigations on reasoning about changes based on rough sets and granular computing have been initiated [19].

There are some important consequences of our considerations for research on approximate reasoning about vague concepts. It is not possible to base such reasoning only on *static* models of the concepts (i.e., approximations of given concepts [9] or membership functions [28] induced from a sample available at a given moment) and on multi-valued logics widely used for reasoning about rough sets or fuzzy sets (see, e.g., [1, 13, 12, 28, 8, 29]). Instead of this we need evolving systems of logics that in open and changing environments will make it possible to gradually acquire knowledge about approximated concepts and reason about them. Along this line an important research perspective arises. Among interesting topics are strategies for modeling of networks supporting such approximate reasoning (e.g., AR schemes and networks [18, 19] can be considered as a step toward developing such strategies), strategies for adaptive revision of such networks, foundations for autonomous systems based on vague concepts.

6 Conclusions

There are two conclusions from our discussion on evolution of approximation spaces in rough set theory. The first one is related to recognizing the importance of inclusion function, generalized in rough mereology to rough inclusion (see, e.g., [14, 15]). This has been used in investigations of information granule calculi, in particular those based on rough mereology (see, e.g., [18, 14, 15]) and approximation spaces based on information granules (see, e.g., [20]). The second one is based on the observation that vague concepts can not be approximated with the high quality by any *static* constructs such as induced membership inclusion functions, approximations or models derived from a partial knowledge about the concepts (e.g., from a sample). Understanding of vague concepts can be only realized in a process in which the induced models are adaptively matching the concepts in the dynamically changing environment. This conclusion seems to have important consequences for understanding vague concepts in soft computing, in particular for the further development of rough set theory in combination with fuzzy sets and other soft computing paradigms.

Acknowledgements

The research has been supported by the grants 3 T11C 002 26 from Ministry of Scientific Research and Information Technology of the Republic of Poland.

References

[1] M. Banerjee, M.K. Chakraborty (2004). Algebras from Rough Sets. In Pal, S.K., Polkowski, L., Skowron, A. (eds.), *Rough-Neural Computing: Techniques for Computing with Words*, Cognitive Technologies, 157-188, Springer-Verlag, Heidelberg, Germany, 2004.

[2] G. Cattaneo (1998). Abstract Approximation Apaces for Rough Theories. In L. Polkowski, A. Skowron (eds.), *Rough Sets in Knowledge Discovery 1: Methodology and Applications*, Studies in Fuzziness and Soft Computing, volume 18, 59–98, Physica-Verlag, Heidelberg, Germany, 1998.

[3] J. Friedman, T. Hastie, R. Tibshirani (2001). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer-Verlag, Heidelberg, Germany.

[4] E. Kloesgen, J. Żytkow (eds.) (2002). *Handbook of Knowledge Discovery and Data Mining*. Oxford University Press, Oxford, UK, 2002.

[5] T,Y, Lin (1998). The Discovery, Analysis and Representation of Data Dependencies in Databases. In L. Polkowski, A. Skowron (eds.), *Rough*

Sets in Knowledge Discovery 1: Methodology and Applications, Studies in Fuzziness and Soft Computing, Volume 18, 107–121, Physica-Verlag, Heidelberg, Germany, 1998.

[6] J. Łukasiewicz (1913). Die Logischen Grundlagen der Wahrscheinlichkeitsrechnung (1913). In Borkowski, L. (ed.), *Jan Łukasiewicz - Selected Works*, North Holland Publishing Company, Amsterdam, London, Polish Scientific Publishers, Warsaw, Poland, 1970.

[7] R. Keefe, P. Smith (eds.) (1997). *Vagueness: A Reader.* MIT Press, Massachusetts, MA, 1997.

[8] J. Pavelka (1979). On Fuzzy Logic I-III. *Zeit. Math Logik Grund. Math.*, volume 25, 45–52, 119-134, 447-464, 1979.

[9] Z. Pawlak (1991). *Rough Sets: Theoretical Aspects of Reasoning about Data.* System Theory, Knowledge Engineering and Problem Solving, Volume 9, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1991.

[10] Z. Pawlak, S.K.M. Wong, W. Ziarko (1990). Rough Sets: Probabilistic Versus Deterministic Approach. In Gaines, B., Boose, J. (eds.), *Machine Learning and Uncertain Reasoning,* volume 3, 227–242, Academic Press, London, UK, 1990.

[11] J.F. Peters, A, Skowron, P. Synak, S. Ramanna (2003). Rough Sets and Information Granulation. In Bilgic, T., Baets, D., Kaynak, O. (eds.), *Tenth International Fuzzy Systems Association World Congress IFSA*, Istanbul, Turkey, June 30-July 2, 2003, Lecture Notes in Artificial Intelligence, volume 2715, 370–377, Springer-Verlag, Heidelberg, Germany, 2003.

[12] L. Polkowski (2002). *Rough Sets: Mathematical Foundations.* Physica-Verlag, Heidelberg, Germany, 2002.

[13] L. Polkowski (2004). A Note on 3-valued Rough Logic Accepting Decision Rules. *Fundamenta Informaticae*, volume 61(1), 37–45, 2004.

[14] L. Polkowski, A. Skowron (1996). Rough Mereology: A New Paradigm for Approximate Reasoning. *International Journal of Approximate Reasoning*, volume 15, 333–365, 1996.

[15] L. Polkowski, A. Skowron (2001). Rough Mereological Calculi of Granules: A Rough Set Approach to Computation. *Computational Intelligence*, volume 17, 472–492, 2001.

[16] S. Read (1995). *Thinking about Logic. An Introduction to the Philosophy of Logic.* Oxford University Press, Oxford, New York, UK, 1995.

[17] A. Skowron, J. Stepaniuk (1996). Tolerance Approximation Spaces. *Fundamenta Informaticae*, volume 27, 245–253, 1996.

[18] A. Skowron, J. Stepaniuk (2004). Information Granules and Rough-Neural Computing. In Pal, S.K., Polkowski, L., Skowron, A. (eds.), *Rough-Neural Computing: Techniques for Computing with Words*, Cognitive Technologies, 43–84, Springer-Verlag, Heidelberg, Germany, 2004.

[19] A. Skowron, P. Synak (2004). Complex Patterns, *Fundamenta Informaticae*, volume 60, 351–366, 2004.

[20] A. Skowron, R. Swiniarski, P. Synak (2004). Approximation Spaces and Information Granulation. In Tsumoto, S., Slowinski, R., Komorowski, J., Grzymala-Busse, J. (eds.), *Fourth International Conference on Rough Sets and Current Trends in Computing (RSCTC 2004)*, June 1-5, 2004, Uppsala, Sweden, Lecture Notes in Artificial Intelligence, volume 3066, 114–123, Springer-Verlag, Heidelberg, Germany, 2004.

[21] R. Słowiński, D. Vanderpooten (1997). Smilarity Relation as a Basis for Rough Set Approximations. In Wang, P.P. (ed.), *Advances in Machine Intelligence & Soft-Computing*, 17–33, Bookwrights, Raleigh, NC, 1997.

[22] P. Stone (2000). *Layered Learning in Multi-Agent Systems: A Winning Approach to Robotic Soccer.* The MIT Press, Cambridge, MA, 2000.

[23] R.S. Sutton, A.G. Barto (1998). *Reinforcement Learning: An Introduction.* MIT Press, Cambridge, MA, 1998.

[24] Y.Y. Yao (1998). Generalized Rough Set Models. In L. Polkowski, A. Skowron (eds.), *Rough Sets in Knowledge Discovery 1: Methodology and Applications*, Studies in Fuzziness and Soft Computing, Volume 18, 286–318, Physica-Verlag, Heidelberg, Germany, 1998.

[25] V. Vapnik (1998). *Statistical Learning Theory.* John Wiley & Sons, New York, NY, 1998.

[26] A. Wojna (2001). Constraint Based Incremental Learning of Classification Rules. In W. Ziarko, Y.Y. Yao (eds.), *Second International Conference on Rough Sets and Current Trends in Computing (RSCTC 2001), Banff, Canada, October 16-19, 2001*, Lecture Notes in Artificial Intelligence, volume 2005, 428-435, Springer-Verlag, Heidelberg, Germany, 2001.

[27] W. Ziarko (1993). Variable Precision Rough Set Model. *Journal of Computer and System Sciences*, volume 46, 39–59, 1993.

[28] L.A. Zadeh (1965). Fuzzy sets. *Information and Control*, volume 8, 333–353, 1965.

[29] L.A. Zadeh (1996). Fuzzy Logic = Computing with Words. *IEEE Transactions on Fuzzy Systems*, volume 2, 103–111, 1996.