Approximation Spaces and Information Granulation

Andrzej Skowron[1], Roman Świniarski[2,3], and Piotr Synak[4]

[1] Institute of Mathematics, Warsaw University,
Banacha 2, 02-097 Warsaw, Poland
skowron@mimuw.edu.pl
[2] Institute of Computer Science, Polish Academy of Sciences,
Ordona 21, 01-237 Warsaw, Poland
[3] Department of Mathematical and Computer Sciences,
San Diego State University,
5500 Campanile Drive San Diego, CA 92182, USA
rswiniar@sciences.sdsu.edu
[4] Polish-Japanese Institute of Information Technology,
Koszykowa 86, 02-008 Warsaw, Poland
synak@pjwstk.edu.pl

Abstract. In this paper, we discuss approximation spaces in a granular computing framework. Such approximation spaces generalise the approaches to concept approximation existing in rough set theory. Approximation spaces are constructed as higher level information granules and are obtained as the result of complex modelling. We present illustrative examples of modelling approximation spaces that include approximation spaces for function approximation, inducing concept approximation, and some other information granule approximations. In modelling of such approximation spaces we use an important assumption that not only objects but also more complex information granules involved in approximations are perceived using only partial information about them.

1 Introduction

The rough set approach is based on the concept of approximation space. Approximation spaces for information systems [1] are defined by partitions or coverings defined by attributes of a pattern space. One can distinguish two basic components in approximation spaces: an uncertainty function and an inclusion function [2]. This approach has been generalised to the rough mereological approach (see, e.g., [3–5]). The existing approaches are based on the observation that the objects are perceived via information about them and due to the incomplete information they can be indiscernible. Hence, with each object one can associate an indiscernibility class, called also (indiscernibility) neighbourhood [6]. In the consequence, testing if a given object belongs to a set is substituted by checking a degree to which its neighbourhood is included into the set. Such an approach covers several generalisations of set approximations like those based on the tolerance relation or the variable precision rough set model [7].

J.F. Peters and A. Skowron (Eds.): Transactions on Rough Sets III, LNCS 3400, pp. 175–189, 2005.

In real-life applications approximation spaces are complex information granules that are not given directly with data but they should be discovered from available data and domain knowledge by some searching strategies (see, e.g., [5,8]). In the paper we present a general approach to approximation spaces based on granular computing. We show that the existing approaches to approximations in rough sets are particular cases of our approach. Illustrative examples include approximation spaces with complex neighbourhoods, approximation spaces for function approximation and for inducing concept approximations. Some other aspects of information granule construction, relevant for approximation spaces, are also presented. Furthermore, we discuss one more aspect of approximation spaces based on the observation that the definition of approximations does not depend only on perception of partial information about objects but also of more complex information granules.

The presented approach can be interpreted in a multi-agent setting [5, 9]. Each agent is equipped with its own relational structure and approximation spaces located in input ports. The approximation spaces are used for filtering (approximating) information granules sent by other agents. Such agents are performing operations on approximated information granules and sending the results to other agents, checking relationships between approximated information granules, or using such granules in negotiations with other agents. Parameters of approximation spaces are analogous to weights in classical neuron models. Agents are performing operations on information granules (that approximate concepts) rather than on numbers. This analogy has been used as a starting point for the rough-neural computing paradigm [10] of computing with words [11].

2 Concept Approximation

In this section we consider the problem of concepts approximation over a universe U^∞ (concepts that are subsets of U^∞). We assume that the concepts are perceived only through some subsets of U^∞, called samples. This is a typical situation in machine learning, pattern recognition, and data mining approaches [12–14]. We explain the rough set approach to induction of concept approximations.

Let $U \subseteq U^\infty$ be a finite sample. By Π_U we denote a perception function from $P(U^\infty)$ into $P(U)$ defined by $\Pi_U(C) = C \cap U$ for any concept $C \subseteq U^\infty$. The problem we consider is how to extend the approximations of $\Pi_U(C)$ to approximation of C over U^∞. In the rough set approach the approximation of a concept is defined by means of a so called approximation space $AS = (U, I, \nu)$, where $I : U \to P(U)$ is an uncertainty function such that $x \in I(x)$ for any $x \in U$, and $\nu : P(U) \times P(U) \to [0, 1]$ is a rough inclusion function (for details see Section 4). We show that the problem can be described as searching for an extension $AS_C = (U^\infty, I_C, \nu_C)$ of the approximation space AS, relevant for approximation of C. This makes it necessary to show how to extend the inclusion function ν from U to relevant subsets of U^∞ that are suitable for the approximation of C. Observe (cf. Definition 5) that for the approximation of C it is enough to induce the necessary values of the inclusion function ν_C without knowing the exact value of $I_C(x) \subseteq U^\infty$ for $x \in U^\infty$.

Let AS be a given approximation space for $\Pi_U(C)$ and let us consider a language L in which the neighbourhood $I(x) \subseteq U$ 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 the case of rule based classifiers patterns of the form $pat(x)$ are defined by feature value vectors.

We assume that for any new object $x \in U^\infty \setminus U$ we can obtain (e.g., as a result of sensor measurement) a pattern $pat(x) \in L$ with semantics $\|pat(x)\|_{U^\infty} \subseteq U^\infty$. However, the relationships between information granules over U^∞ like sets: $\|pat(x)\|_{U^\infty}$ and $\|pat(y)\|_{U^\infty}$, for different $x, y \in U^\infty$ (or between $\|pat(x)\|_{U^\infty}$ and $y \in U^\infty$), are, in general, known only if they can be expressed by relationships between the restrictions of these sets to the sample U, i.e., between sets $\Pi_U(\|pat(x)\|_{U^\infty})$ and $\Pi_U(\|pat(y)\|_{U^\infty})$.

The set of patterns $\{pat(x) : x \in U\}$ is usually not relevant for approximation of the concept $C \subseteq U^\infty$. Such patterns are too specific or not general enough, and can directly be applied only to a very limited number of new objects. However, by using some generalisation 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 the case of a rule based classifier one can search for such candidate patterns among sets definable by subsequences of feature value vectors corresponding to objects from the sample U. The set $PATTERNS(AS, L, C)$ can be selected by using some quality measures checked on meanings (semantics) of its elements restricted to the sample U (like the number of examples from the concept $\Pi_U(C)$ and its complement that support a given pattern). Then, on the basis of properties of sets definable by those patterns over U we induce approximate values of the inclusion function ν_C on subsets of U^∞ definable by any of such pattern and the concept C.

Next, we induce the value of ν_C on pairs (X, Y) where $X \subseteq U^\infty$ is definable by a pattern from $\{pat(x) : x \in U^\infty\}$ and $Y \subseteq U^\infty$ is definable by a pattern from $PATTERNS(AS, L, C)$.

Finally, for any object $x \in U^\infty \setminus U$ we induce the approximation of the degree $\nu_C(\|pat(x)\|_{U^\infty}, C)$ applying a conflict resolution strategy $Conflict\_res$ (a voting strategy, in the case of rule based classifiers) to two families of degrees:

$$\{\nu_C(\|pat(x)\|_{U^\infty}, \|pat\|_{U^\infty}) : pat \in PATTERNS(AS, L, C)\}, \tag{1}$$

$$\{\nu_C(\|pat\|_{U^\infty}, C) : pat \in PATTERNS(AS, L, C)\}. \tag{2}$$

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_C – it is enough to induce the values of the inclusion function ν_C. Observe that the defined extension ν_C of ν to some subsets of U^∞ makes it possible to define an approximation of the concept C in a new approximation space AS_C by using Definition 5.

In this way, the rough set approach to induction of concept approximations can be explained as a process of inducing a relevant approximation space.

3 Granule Approximation Spaces

Using the granular computing approach (see, e.g., [5]) one can generalise the approximation operations for sets of objects, known in rough set theory, to arbitrary information granules. The basic concept is the rough inclusion function ν [3–5].

First, let us recall the definition of an information granule system [5].

Definition 1. *An information granule system is any tuple* $GS = (E, O, G, \nu)$ *where* E *is a set of elements called elementary information granules,* O *is a set of (partial) operations on information granules,* G *is a set of information granules constructed from* E *using operations from* O*, and* $\nu : G \times G \longrightarrow [0, 1]$ *is a (partial) function called rough inclusion.*

The main interpretation of rough inclusion is to measure the inclusion degree of one granule in another.

In the paper we use the following notation: $\nu_p(g, g')$ denotes that $\nu(g, g') \geq p$ holds; $Gran(GS) = G$; $\mathcal{G}$ denotes a given family of granule systems. For every non-empty set X, let $P(X)$ denote the set of all subsets of X.

We begin with the general definition of approximation space in the context of a family of information granule systems.

Definition 2. *Let* $\mathcal{G}$ *be a family of information granule systems. A granule approximation space with respect to* $\mathcal{G}$ *is any tuple*

$$\mathcal{AS}_{\mathcal{G}} = (GS, G, Tr), \tag{3}$$

where GS *is a selected (initial) granule system from* $\mathcal{G}$*;* $G \subseteq Gran(GS)$ *is some collection of granules; a transition relation* Tr *is a binary relation on information granule systems from* $\mathcal{G}$*, i.e.,* $Tr \subseteq \mathcal{G} \times \mathcal{G}$*.*

Let $GS \in \mathcal{G}$. By $Tr[GS]$ we denote the set of all information granule systems reachable from GS:

$$Tr[GS] = \{GS' \in \mathcal{G} : GS\ Tr^*\ GS'\}, \tag{4}$$

where Tr^* is the reflexive and transitive closure of the relation Tr. By $Tr[GS, G]$ we denote the set of all Tr-terminal granule systems reachable from GS that consist of information granules from G:

$$Tr[GS, G] = \\ \{GS' : (GS, GS') \in Tr^* \text{ and } G \subseteq Gran(GS') \text{ and } Tr[GS'] = \{GS'\}\}. \tag{5}$$

The elements of $Tr[GS, G]$ are called the candidate granule systems for approximation of information granules from G, generated by Tr from GS (G-candidates, for short). For any system $GS^* \in Tr[GS, G]$ we define approximations of granules from G by information granules from $Gran(GS^*) \setminus G$. Searching for granule systems from $Tr[GS, G]$ that are relevant for approximation of given information granules is one of the most important tasks in granular computing.

By using granule approximation space $\mathcal{AS}_{\mathcal{G}} = (GS, G, Tr)$, for a family of granule systems $\mathcal{G}$, we can define approximation of a given granule $g \in G$ in terms of its lower and upper approximations[1]. We assume that there is additionally

[1] If there is no contradiction we use $\mathcal{AS}$ instead of $\mathcal{AS}_{\mathcal{G}}$.

given a "make granule" operation $\oplus : P(G^*) \longrightarrow G^*$, where $G^* = Gran(GS^*)$, for any $GS^* \in Tr(GS, G)$, that constructs a single granule from a set of granules. A typical example of $\oplus$ is set theoretical union, however, it can be also realised by a complex classifier. The granule approximation is thus defined as follows:

Definition 3. *Let $0 \leq p < q \leq 1$, $\mathcal{AS} = (GS, G, Tr)$ be a granule approximation space, and $GS^* \in Tr[GS, G]$. The $(\mathcal{AS}, GS^*, \oplus, q)$-lower approximation of $g \in G$ is defined by*

$$LOW(\mathcal{AS}, GS^*, \oplus, q)(g) = \oplus(\{g' \in Gran(GS^*) \setminus G : \nu^*(g', g) \geq q\}) \quad (6)$$

where ν^* denotes the rough inclusion of GS^*.

The $(\mathcal{AS}, GS^*, \oplus, p)$-upper approximation of g is defined by

$$UPP(\mathcal{AS}, GS^*, \oplus, p)(g) = \oplus(\{g' \in Gran(GS^*) \setminus G : \nu^*(g', g) > p\}) \quad (7)$$

where ν^* denotes the rough inclusion of GS^*.

The numbers p, q can be interpreted as inclusion degrees that make it possible to control the size of the lower and upper approximations. In the case when $p = 0$ and $q = 1$ we have the case of full inclusion (lower approximation) and any non-zero inclusion (upper approximation).

One can search for optimal approximations of granules from G defined by $GS^* \in Tr[GS, G]$ using some optimisation criteria or one can search for relevant fusion of approximations of granules from G defined by $GS^* \in Tr[GS, G]$.[2]

In the following sections we discuss illustrative examples showing that the above scheme generalises several approaches to approximation spaces and set approximations. In particular, we include examples of information granules G and their structures, the rough inclusion ν as well as the $\oplus$ operation.

4 Approximation Spaces

Let us recall the definition of an approximation space from [1, 2]. For simplicity of reasoning we omit parameters that label components of approximation spaces.

Definition 4. *An approximation space is a system $AS = (U, I, \nu)$, where*

- *U is a non-empty finite set of objects,*
- *$I : U \rightarrow P(U)$ is an uncertainty function such that $x \in I(x)$ for any $x \in U$,*
- *$\nu : P(U) \times P(U) \rightarrow [0, 1]$ 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 standard rough inclusion function ν_{SRI} defines the degree of inclusion between two subsets of U by

$$\nu_{SRI}(X, Y) = \begin{cases} \frac{card(X \cap Y)}{card(X)} & \text{if } X \neq \emptyset \\ 1 & \text{if } X = \emptyset. \end{cases} \quad (8)$$

[2] This problem will be investigated elsewhere.

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

Definition 5. *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*

$$LOW_q\left(AS, X\right) = \left\{x \in U : \nu\left(I\left(x\right), X\right) \geq q\right\}, \tag{9}$$

$$UPP_p\left(AS, X\right) = \left\{x \in U : \nu\left(I\left(x\right), X\right) > p\right\}, \tag{10}$$

respectively.

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 generalises several approaches existing in the literature like those based on equivalence or tolerance indiscernibility relation as well as those based on exact inclusion of indiscernibility classes into concepts [1, 7].

Let us observe that the above definition of approximations is a special case of Definition 3. The granule approximation space $\mathcal{AS} = (GS, G, Tr)$ corresponding to $AS = (U, I, \nu)$ can be defined by

1. GS consisting of information granules being subsets of U (in particular, containing neighbourhoods that are values of the uncertainty function I) and of rough inclusion $\nu = \nu_{SRI}$.
2. $G = P(U)$.
3. $Tr[GS, G]$ consisting of exactly two systems: GS and GS^* such that
   - $Gran(GS^*) = G \cup \{(x, I(x)) : x \in U\}$;
   - the rough inclusion ν is extended by $\nu((x, I(x)), X) = \nu(I(x), X)$ for $x \in U$ and $X \subseteq U$.

Suppose the "make granule" operation $\oplus$ is defined by

$$\oplus(\{(x, \cdot) : x \in Z\}) = Z \text{ for any } Z \subseteq U.$$

Then for the approximation space AS and granule approximation space $\mathcal{AS}$ we have the following:

Proposition 1. *Let $0 \leq p < q \leq 1$. For any $X \in P(U)$ we have:*

$$LOW_q\left(AS, X\right) = LOW(\mathcal{AS}, GS^*, \oplus, q)(X), \tag{11}$$

$$UPP_p\left(AS, X\right) = UPP(\mathcal{AS}, GS^*, \oplus, p)(X). \tag{12}$$

5 Approximation Spaces with Complex Neighbourhoods

In this section we present approximation spaces that have more complex uncertainty functions. Such functions define complex neighbourhoods of objects, e.g., families of sets. This aspect is very important from the point of view of complex concepts approximation. A special case of complex uncertainty functions is such with values in $P^2(U)$, i.e., $I : U \longrightarrow P^2(U)$. Such uncertainty functions appear, e.g., in the case of the similarity based rough set approach. One can define $I(x)$ to be a family of all maximal cliques defined by the similarity relation which x belongs to (for details see Section 8).

We obtain the following definition of approximation space:

Definition 6. *A k-th order approximation space is any tuple $AS = \left(U, I^k, \nu\right)$, where*

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

Let us note that up to the above definition there can be given different uncertainty functions for different levels of granulation. The inclusion function can be also defined in this way, however, in most cases we induce it recursively from ν. For example, in the case of set approximation by means of given approximation space $AS = \left(U, I^k, \nu\right)$ we are interested in an inclusion function $\nu^k : P^k(U) \times P(U) \to [0,1]$, defined recursively by the corresponding relation ν_p^k as follows

- $\nu_p^k(\mathcal{Y}, X)$ iff $\exists Y \in \mathcal{Y}\ \ \nu_p^{k-1}(Y, X)$,
- $\nu_p^1(Y, X)$ iff $\nu_p(Y, X)$,

for $X \subseteq U$ and $\mathcal{Y} \subseteq P^k(U)$.

The definition of set approximations for the k-th order approximation spaces depends on the way the values of uncertainty function are perceived. To illustrate this point of view we consider the following two examples: the complete perception of neighbourhoods and the perception defined by the intersection of the family $I(x)$. In the former case we consider a new definition of set approximations.

Definition 7. *Let $0 \leq p < q \leq 1$. For any k-th order approximation space $AS = \left(U, I^k, \nu\right)$, ν^k induced from ν, and any subset $X \subseteq U$ the q-lower and the p-upper approximation of X in AS are defined by*

$$LOW_q\left(AS, X\right) = \left\{x \in U : \nu^k\left(I^k(x), X\right) \geq q\right\}, \tag{13}$$

$$UPP_p\left(AS, X\right) = \left\{x \in U : \nu^k\left(I^k(x), X\right) > p\right\}, \tag{14}$$

respectively.

We can observe, that the approximation operations for those two cases are, in general, different.

Proposition 2. *Let us denote by $AS^\cap = (U, I^\cap, \nu)$ the approximation space obtained from the second order approximation space $AS = \left(U, I^2, \nu\right)$ assuming $I^\cap(x) = \cap I^2(x)$ for $x \in U$. We also assume that $x \in Y$ for any $Y \in I^2(x)$ and $x \in U$. Then, for $X \subseteq U$ we have*

$$LOW\left(AS, X\right) \subseteq LOW\left(AS^\cap, X\right) \subseteq X \subseteq UPP\left(AS^\cap, X\right) \subseteq UPP\left(AS, X\right). \tag{15}$$

One can check (in an analogous way as in Section 4) that the above definition of approximations is a special case of Definition 3.

6 Relation and 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. We can consider approximation of R by an approximation space $AS = (U \times U, I, \nu)$ in an analogous way as in Definition 5:

$$LOW_q(AS, R) = \{(x, y) \in U \times U : \nu(I(x, y)), X) \geq q\}, \tag{16}$$

$$UPP_p(AS, R) = \{(x, y) \in U \times U : \nu(I(x, y), X) > p\}, \tag{17}$$

for $0 \leq p < q \leq 1$. This definition can be also easily extended to the case of complex uncertainty function as in Definition 7. However, the main problem is how to construct relevant approximation spaces, i.e., how to define uncertainty and inclusion functions. One of the solutions is the following uncertainty function

$$I(x, y) = I(x) \times I(y), \tag{18}$$

(assuming that one dimensional uncertainty function is given) and the standard inclusion, i.e., $\nu = \nu_{SRI}$.

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 $\{(x, f(x)) : x \in Dom\}$. We can easily see that if we apply the above definition of relation approximation to f (it 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 different way, e.g., by extending the uncertainty function as well as the inclusion function on subsets of $U \times U$. We assume that the value $I^*(x, y)$ of the uncertainly function, called the neighbourhood (or the window) of (x, y), for $(x, y) \in U \times U$, is defined by

$$I^*(x, y) = I(x) \times I(y). \tag{19}$$

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

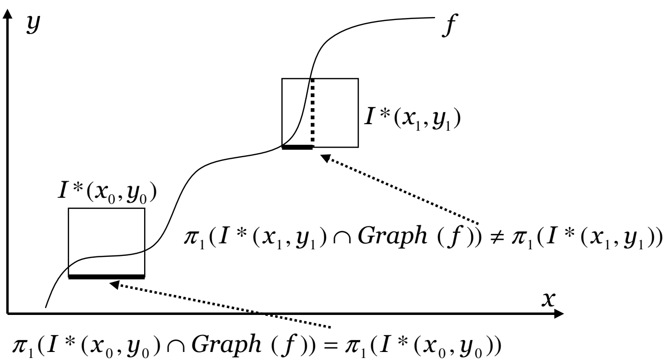

Fig. 1. Function approximation.

$I^*(x,y) \cap Graph(f)$ is included into $Graph(f)$. If $I(x) \cap Dom \neq \emptyset$, one can consider a ratio r of the fluctuation in $I(y)$ of the function $f \upharpoonright I(x)$ to the fluctuation in $I(x)$, where by $f \upharpoonright I(x)$ we denote the restriction of the function f to $I(x)$. If $r = 1$ then the window is in the lower approximation of $Graph(f)$; if $0 < r \leq 1$ then the window $I^*(x,y)$ is in the upper approximation of $Graph(f)$. If $I(x) \cap Dom = \emptyset$ then the degree r is equal to zero. 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 $I(x)$. This can be formally defined by the following condition:

$$\pi_1(I^*(x,y) \cap Graph(f)) = \pi_1(I^*(x,y)) \tag{20}$$

where π_1 denotes the projection on the first coordinate. Condition (20) is equivalent to:

$$\forall x' \in I(x)\ \exists y' \in I(y)\ y' = f(x'). \tag{21}$$

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

$$\nu^*(X,Y) = \begin{cases} \frac{card(\pi_1(X \cap Y))}{card(\pi_1(X))} & \text{if } \pi_1(X) \neq \emptyset \\ 1 & \text{if } \pi_1(X) = \emptyset. \end{cases} \tag{22}$$

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

We have the following proposition:

Proposition 3. *Let $AS^* = (U \times U, I^*, \nu^*)$ be an approximation space with I^*, ν^* defined by (19), (22), respectively, and let $f : Dom \longrightarrow U$ where $Dom \subseteq U$. Then we have*

1. $(x,y) \in LOW_1(AS^*, Graph(f))$ *if and only if* $\forall x' \in I(x)\ \exists y' \in I(y)\ y' = f(x')$;
2. $(x,y) \in UPP_0(AS^*, Graph(f))$ *if and only if* $\exists x' \in I(x)\ \exists y' \in I(y)\ y' = f(x')$.

In the case of arbitrary parameters p, q satisfying $0 \leq p < q \leq 1$ we have

Proposition 4. *Let $AS^* = (U \times U, I^*, \nu^*)$ be an approximation space with I^*, ν^* defined by* (19), (22), *respectively, and let $f : Dom \longrightarrow U$ where $Dom \subseteq U$. Then we have*

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

In our example, we define the inclusion degree between two subsets of Cartesian product using, in a sense, the inclusion degree between their projections. Hence, subsets of Cartesian products are perceived by projections.

Again, one can consider the definition of approximation space for function approximation as a special case of the granule approximation space introduced in Definition 2 with the non standard rough inclusion introduced in this section.

7 Concept Approximation by Granule Approximation Space

The granule approximation space $\mathcal{AS} = (GS, G, Tr)$ modelling the described process of concept approximations under fixed U^∞, $C \subseteq U^\infty$, sets of formulas (patterns) $\{pat(x) : x \in U\}$, $PATTERNS(AS, L, C)$ and their semantics $\|\cdot\|_{U^\infty}$ can be defined by

1. GS consisting of the following granules: $C \in P(U^\infty)$, the sample $U \subseteq U^\infty$, $C \cap U$, $U \setminus C$, sets $\|pat(x)\|_U$, defined by $pat(x)$ for any $x \in U$, and the rough inclusion $\nu = \nu_{SRI}$.
2. $G = \{C\}$.
3. The transition relation Tr extending GS to GS' and GS' to GS^*. $Gran(GS')$ is extended from $Gran(GS)$ by the following information granules: the sets $\|pat(x)\|_{U^\infty}$, defined by $pat(x)$ for any $x \in U^\infty$, sets $\|pat\|_{U^\infty}$, for $pat \in PATTERNS(AS, L, C)$. The rough inclusion is extended using estimations described above. GS^* is constructed as follows:
   - $Gran(GS^*) = G \cup$ $\cup\{(x, \|pat(x)\|_{U^\infty}, \|pat\|_{U^\infty}) : x \in U^\infty \wedge pat \in PATTERNS(AS, L, C)\}$;
   - The rough inclusion ν is extended by:

$$\nu((x, X, Y), C) =$$
$$Conflict\_res(\{\nu_C(X, Y) : Y \in \mathcal{Y}\}, \{\nu_C(Y, C) : Y \in \mathcal{Y}\}) \quad (23)$$

     where $X, Y \subseteq U^\infty$, $\mathcal{Y} \subseteq P(U^\infty)$ are sets and the family of sets on which values of ν_C have been estimated in (1) and (2);
   - The operation "make granule" $\oplus$ satisfies the following constraint:

$$\oplus\{(x, \cdot, \cdot) : x \in Z\} = Z \text{ for any } Z \subseteq C^\infty.$$

8 Relational Structure Granulation

In this section we discuss an important role that the relational structure granulation [5], [8] plays in searching for relevant patterns in approximate reasoning, e.g., in searching for relevant approximation patterns (see Section 2 and Figure 2).

Let us recall, that the uncertainty (neighbourhood) function of an approximation space forms basic granules of knowledge about the universe U. Let us consider the case where the values of neighbourhood function are from $P^2(U)$.

Assume that together with an information system $\mathbb{A} = (U, A)$ [1] there is also given a similarity relation τ defined on vectors of attribute values. This relation can be extended to objects. An object $y \in U$ is similar to a given object $x \in U$ if the attribute value vector on x is τ-similar to the attribute value vector on y. Now, consider a neighbourhood function defined by $I_{\mathbb{A},\tau}(x) = \{[y]_{IND(A)} : x\tau y\}$.

Neighbourhood functions cause a necessity of further granulation. Let us consider granulation of a relational structure R by neighbourhood functions. We would like to show that due to the relational structure granulation we obtain new information granules of more complex structure and, in the consequence, more general neighbourhood functions than those discussed above. Hence, basic granules of knowledge about the universe corresponding to objects become more complex.

Assume that a relational structure R and a neighbourhood function I are given. The aim is to define a new relational structure R_I called the I–granulation[3] of R. This is done by granulation of all components of R, i.e., relations and functions (see Section 6), by means of I.

The relational structure granulation defines new patterns that are created for pairs of objects together with some inclusion and closeness measures defined among them. Such patterns can be used for approximation of a target concept (or a concept on an intermediate level of hierarchical construction) over objects composed from pairs (x, y) interpreted, e.g., as parts of some more compound objects. Such compound objects are often called structured or complex objects.

Certainly, to induce approximations of high quality it is necessary to search for relevant patterns for concept approximation expressible in a given language. This problem, known as feature selection, is widely discussed in machine learning, pattern recognition, and data mining (see, e.g., [12–14]).

Let us consider an exemplary degree structure $D = ([0, 1], \leq)$ and its granulation $D_{I_0} = (P([0, 1]), \leq_{I_0})$ by means of an uncertainty function $I_0 : [0, 1] \rightarrow P([0, 1])$ defined by $I_0(x) = \{y \in [0, 1] : [10^k x] = [10^k y]\}$, for some integer k, where for $X, Y \subseteq [0, 1]$ we assume $X \leq_{I_0} Y$ iff $\forall x \in X, \forall y \in Y\ x \leq y$. Let $\{X_s, X_m, X_l\}$ be a partition of $[0, 1]$ satisfying $x < y < z$ for any $x \in X_s$, $y \in X_m$, $z \in X_l$. Let $AS_0 = ([0, 1], I_0, \nu)$ be an approximation space with the standard inclusion function ν. We denote by S, M, L the lower approximations of X_s, X_m, X_l in AS_0, respectively, and by S–M, M–L the boundary regions between X_s, X_m and X_m, X_l, respectively. Moreover, we assume $S, M, L \neq \emptyset$. In this way we obtain restriction of D_{I_0} to the structure $(Deg, \leq_{I_0})$, where $Deg = \{S, S\text{–}M, M, M\text{–}L, L\}$. Now, for a given (multi-sorted) structure $(U, P(U), [0, 1], \leq, I_0, \nu)$, where $\nu : P(U) \times P(U) \rightarrow [0, 1]$ is an inclusion function, we can define its I_0–granulation by

$$(U, P(U), Deg, \leq_{I_0}, \{\nu_d\}_{d \in Deg}) \tag{24}$$

where $Deg = \{S, S\text{–}M, M, M\text{–}L, L\}$ and $\nu_d(X, Y)$ iff $\nu_p(X, Y)$ for some p, d', such that $p \in d'$ and $d \leq_{I_0} d'$.

[3] In general, granulation is defined using the uncertainty function and the inclusion function from a given approximation space AS. For simplicity, we restrict our initial examples to I–granulation only.

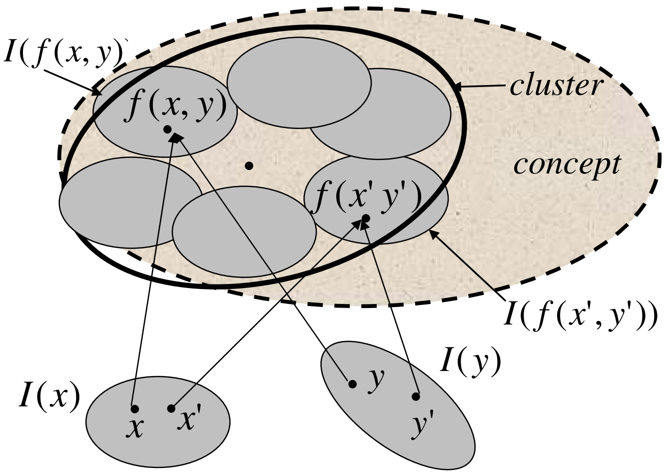

Fig. 2. Relational structure granulation.

Thus, for any object there is defined a neighbourhood specified by the value of uncertainty function from an approximation space. From those neighbourhoods some more relevant ones (e.g., for the considered concept approximation), should be discovered. Such neighbourhoods can be extracted by searching in a space of neighbourhoods generated from values of uncertainty function by applying to them some operations like generalisation operations, set theoretical operations (union, intersection), clustering, and operations on neighbourhoods defined by functions and relations in an underlying relational structure[4]. Figure 2 illustrates an exemplary scheme of searching for neighbourhoods (patterns, clusters) relevant for concept approximation. In this 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 neighbourhoods 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 neighbourhoods $\mathcal{F} = \{I(f(x',y')) : (x',y') \in I(x) \times I(y)\}$. From this family $\mathcal{F}$ a subfamily $\mathcal{F}'$ of neighbourhoods can be chosen that consists of neighbourhoods with some properties relevant for approximation (e.g., neighbourhoods with sufficiently large support and/or confidence with respect to a given target concept). Next, the subfamily $\mathcal{F}'$ can be generalised to clusters that are relevant for the concept approximation, i.e., clusters sufficiently included into the approximated concept (see Figure 2). Observe also that the inclusion degrees can be measured by granulation of the inclusion function from the relational structure.

Now, let us present some examples illustrating information system granulation on searching for concept approximation.

Let $\mathbb{A} = (U, A)$ be an information system with universe U of objects described by some features from an attribute set A. In many cases, there are also given

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

some additional relational structures on V_a, e.g., relations r_a defined for each attribute $a \in A$. Using $\{r_a\}_{a \in A}$ one can define relational structures over U in many different ways. For example, $r_a \subseteq V_a \times V_a$ can be a similarity relation for any $a \in A$. Such relations can be used to define similarity between objects $Sim \subseteq U \times U$, e.g., by $xSimy$ iff $r_a(a(x), a(y))$ for any $a \in A$. Then, for each $x \in U$ one can consider a relational structure R_x defined by a tolerance class $Sim(x) = \{y \in U : xSimy\}$ with relation Sim reduced to $Sim(x)$. In this way we obtain a new universe $U_{Rel} = \{R_x : x \in U\}$.

The trajectories of objects in time, $o(t)$, are basic objects in spatio-temporal reasoning and time series analysis. By restricting $o(t)$ to some time window of fixed length one can construct the basic relational structures forming objects of a new information system.

In the case of decision problems, i.e., when the initial information system $\mathbb{A}$ is a decision table, the task is to define relevant decisions for the considered problems and to define conditions making it possible to approximate new decision classes. These new decisions can be related to different tasks, e.g., to prediction in time series [13, 14], decision class approximations (robust with respect to deviations defined by similarities) [13, 14], and preference analysis [15]. For solving such tasks, the methods searching for relevant granulation of relational structures representing objects are very important.

Relational structures also arise in many pattern recognition problems as the result of (perception) representation of the object structure or data dimension reduction. Information granules considered in such applications are equal to elementary granules (indiscernibility classes) of information systems determined by some relational structures. Below we discuss this kind of granulation in more detail.

Let $\mathbb{A} = (U, A)$ be an information system where $a : U \longrightarrow V_a$ for any $a \in A$. Assume that $f : X \longrightarrow U$ and $b_a(x) = a(f(x))$ for any $x \in X$ and $a \in A$. Any such pair $(\mathbb{A}, f)$ defines a relational structure $\mathcal{R}$ on $Y = X \cup U \cup \bigcup_{a \in A} V_a$ with unary relations r_U, r_X, r_{V_a} and binary relations r_f and r_a, for $a \in A$, where $y \in r_X$ iff $y \in X$, $y \in r_U$ iff $y \in U$, $y \in r_{V_a}$ iff $y \in V_a$, for any $y \in Y$ and $a \in A$; $r_f \subseteq Y \times Y$ is a partial function defined by f, $r_a \subseteq Y \times Y$ is a partial function defined by a for any $a \in A$. Information granules over such a relational structure $\mathcal{R}$ are B-indiscernibility classes (elementary granules) of the information system $\mathcal{B} = (X, B)$ where $B = \{b_a : a \in A\}$. Elementary granules $[x]_{IND(B)}$ for $x \in X$, where $IND(B)$ is the B-indiscernibility relation, have the following property:

$$\begin{aligned} [x]_{IND(B)} &= f^{-1}(Inf_A^{-1}(Inf_A(f(x)))) \\ &= \bigcup \{y \in [f(x)]_{IND(A)} : \ f^{-1}(\{y\})\} \end{aligned} \tag{25}$$

where $f^{-1}(Z)$ denotes the counter-image of the set Z with respect to the function f.

The function f is used in pattern recognition [12, 14] applications to extract relevant parts of classified objects or to reduce the data dimension. Searching for relevant (with respect to some target concepts) granules of the form defined in (25) is performed by tuning f and A.

9 Conclusions

We discussed the problems of approximation space modelling for concept approximation. We presented consequences of the assumption that information granules involved in concept approximations are perceived by partial information about them. Illustrative examples of approximation spaces were also included. We emphasised the role of relational structure granulation in searching for relevant approximation spaces.

In our further work we would like to use the presented approach for modelling of searching processes for relevant approximation spaces using data and domain knowledge represented, e.g., in a natural language.

Acknowledgements

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

References

1. Pawlak, Z.: Rough Sets: Theoretical Aspects of Reasoning about Data. Volume 9 of System Theory, Knowledge Engineering and Problem Solving. Kluwer Academic Publishers, Dordrecht, The Netherlands (1991)
2. Skowron, A., Stepaniuk, J.: Tolerance approximation spaces. Fundamenta Informaticae **27** (1996) 245–253
3. Polkowski, L., Skowron, A.: Rough mereology: A new paradigm for approximate reasoning. International Journal of Approximate Reasoning **15** (1996) 333–365
4. Polkowski, L., Skowron, A.: Towards adaptive calculus of granules. In Zadeh, L.A., Kacprzyk, J., eds.: Computing with Words in Information/Intelligent Systems. Physica-Verlag, Heidelberg, Germany (1999) 201–227
5. Skowron, A., Stepaniuk, J.: Information granules and rough-neural computing. In [10] 43–84
6. Lin, T.Y.: The discovery, analysis and representation of data dependencies in databases. In Polkowski, L., Skowron, A., eds.: Rough Sets in Knowledge Discovery 1: Methodology and Applications. Volume 18 of Studies in Fuzziness and Soft Computing. Physica-Verlag, Heidelberg, Germany (1998) 107–121
7. Ziarko, W.: Variable precision rough set model. Journal of Computer and System Sciences **46** (1993) 39–59
8. Peters, J.F., Skowron, A., Synak, P., Ramanna, S.: 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. Volume 2715 of Lecture Notes in Artificial Intelligence. Springer-Verlag, Heidelberg, Germany (2003) 370–377
9. Skowron, A.: Toward intelligent systems: Calculi of information granules. Bulletin of the International Rough Set Society **5** (2001) 9–30
10. Pal, S.K., Polkowski, L., Skowron, A., eds.: Rough-Neural Computing: Techniques for Computing with Words. Cognitive Technologies. Springer-Verlag, Heidelberg, Germany (2003)

11. Zadeh, L.A.: A new direction in AI: Toward a computational theory of perceptions. AI Magazine **22** (2001) 73–84
12. Duda, R.O., Hart, P.E., Stork, D.G.: Pattern Classification. John Wiley & Sons, New York, NY (2001)
13. Hastie, T., Tibshirani, R., Friedman, J.: The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer-Verlag, Heidelberg, Germany (2001)
14. Kloesgen, W., Żytkow, J., eds.: Handbook of Knowledge Discovery and Data Mining. Oxford University Press, New York, NY (2002)
15. Słowiński, R., Greco, S., Matarazzo, B.: Rough set analysis of preference-ordered data. In Alpigini, J.J., Peters, J.F., Skowron, A., Zhong, N., eds.: Third International Conference on Rough Sets and Current Trends in Computing RSCTC, Malvern, PA, October 14-16, 2002. Volume 2475 of Lecture Notes in Computer Science. Springer-Verlag, Heidelberg, Germany (2002) 44–59