HIERARCHY OF INFORMATION GRANULES

Andrzej Skowron[1] Jaroslaw Stepaniuk[2]
[1] Institute of Mathematics, Warsaw University,
Banacha 2, 02-097 Warsaw, Poland,
E-mail: skowron@mimuw.edu.pl
[2] Institute of Computer Science,
Bialystok University of Technology,
Wiejska 45A, 15-351 Bialystok, Poland,
E-mail: jstepan@ii.pb.bialystok.pl

Abstract

In this paper we generalize a simple notion of elementary granules represented by attribute value vectors to the case of hierarchical granules. We claim that in many application areas related to knowledge discovery and data mining there is a need for algorithmic methods to discover much more complex information granules and relations between them than investigated so far. We discuss examples of information granules and we consider two kinds of basic relations between them, namely inclusion and closeness. The relations between more complex information granules can be defined by extension of the relations defined on parts of the information granules.

Keywords

rough set theory, tolerance rough sets, information granulation, knowledge discovery in databases

1 Introduction

Last years saw a rapid growth of interest in so-called granular computing. It is geared toward representing and processing basic chunks of information. A form of granular computing was introduced by Zadeh. Granular computing has grown out of studies of fuzzy sets and rough sets. Zadeh defines information granulation as follows:
"Information granulation involves partitioning a class of objects (points) into granules, with a granule being a clump of objects (points) which are drawn together by indistinguishability, similarity or functionality." Zadeh also introduced the following terms related to information granules: "computing with words", "granular computing" and "from measurements to perception". We follow a rough set way of constructing information granules.

The theory of rough sets provides a powerful foundation for discovery of important regularities in data and for objects classification. In recent years numerous successful applications of rough set methods for real-life data have been developed (see e.g. [7], [8], [6]).

We will now describe in some detail rough set models. Rough set approach has been used in a lot of applications aimed to description of concepts. In most cases only approximate descriptions of concepts can be constructed because of incomplete information about them. Let us consider a typical example for classical rough set approach when concepts are described by positive and negative examples. In such situations it is not always possible describe concepts exactly, since some positive and negative examples of the concepts being described inherently can not be distinguished one from another. Rough set theory was proposed [7] as a new approach to vague concept description from incomplete data. The rough set approach to processing of incomplete data is based on the lower and the upper approximation. The rough set is defined as the pair of two crisp sets corresponding to approximations. If both approximations of a given subset of the universe are exactly the same, then one can say that the subset mentioned above is definable with respect to available information. Otherwise, one can consider it as roughly definable. Suppose we are given a finite non-empty set U of objects, called the universe. Each object of U is characterized by a description constructed, for example from a set of attribute values. In standard rough set approach [7] introduced by Pawlak an equivalence relation (reflexive, symmetric and transitive relation) on the universe of objects is defined from equivalence relations on the attribute values. In particular, this equivalence relation is constructed assuming the existence of the equality relation on attribute values. Two different objects are indiscernible in view of available information, because with these objects the same information can be associated. Thus, information associated with objects from the universe generates an indiscernibility relation in this universe. In the standard rough set model the lower approximation of any subset $X \subseteq U$ is defined as the union of all equivalence classes fully included in X. On the other hand the upper approximation of X is defined as the union of all equivalence classes with a non-empty intersection with X.

In real data sets usually there is some noise, caused for example from imprecise measurements or mistakes made during collecting data. In such situations the notions of "full inclusion" and "non-empty intersection" used in approximations definition are too restrictive. Some extensions in this direction have been proposed in the variable precision rough set model.

One of the problems we are interested in is the following: given a subset $X \subseteq U$ or a relation $R \subseteq U \times U$, define X or R in terms of the available information. We discuss an approach based on generalized approximation spaces introduced and investigated in [9]. We combine in one model not only some extension of an indiscernibility relation but also some extension of the standard inclusion used in definitions of approximations in the standard rough set model.

There are several modifications of the original approximation space definition [7]. The first one concerns the so called uncertainty function. Information about an object, say x is represented for example by its attribute value vector. Let us denote the set of all objects with similar (to attribute value vector of x) value vectors by $I(x)$. In the standard rough set approach [7] all objects with the same value vector create the indiscernibility class. The relation $y \in I(x)$ is in this case an equivalence relation. The second modification of the approximation space definition introduces a generalization of the rough membership function. We assume that to answer a question whether an object x belongs to an object set X we have to answer a question whether $I(x)$ is in some sense included in X.

In the presented paper we generalize a simple notion of elementary granules represented by attribute value vectors as well as closeness (inclusion) relation to the case of hierarchical granules representing concepts. We claim that in many application areas related to knowledge discovery and data mining there is a need for algorithmic methods to discover much more complex information granules and relations between them than investigated so far. We discuss examples of information granules and we consider two kinds of basic relations between them, namely (rough) inclusion and closeness. The relations between more complex information granules can be defined by extension of the relations defined on parts of the information granules.

2 Information Granules and Tolerance Rough Sets

We present general definition of an approximation space [9], [11] which can be used for example for introducing the tolerance based rough set model and the variable precision rough set model.

For every non-empty set U, let $P(U)$ denote the set of all subsets of U.

Definition 1 *A parameterized approximation space is a system* $AS_{\#,\$} = (U, I_{\#}, \nu_{\$})$, *where*

- U *is a non-empty set of objects,*
- $I_{\#} : U \to P(U)$ *is an uncertainty function,*
- $\nu_{\$} : P(U) \times P(U) \to [0,1]$ *is a rough inclusion function,*

and $\#, \$$ *are denoting vectors of parameters.*

The uncertainty function defines for every object x a set of similarly described objects (elementary granule). A constructive definition of uncertainty function can be based on the assumption that some metrics (distances) are given on attribute values. For example, if for some attribute $a \in A$ a metric $\delta_a : V_a \times V_a \longrightarrow [0,\infty)$ is given, where V_a is the set of all values of attribute a then one can define the following uncertainty function:

$$y \in I_a^{f_a}(x) \text{ if and only if } \delta_a(a(x), a(y)) \leq f_a(a(x), a(y)),$$

where $f_a : V_a \times V_a \to [0, \infty)$ is a given threshold 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 rough inclusion function defines the degree of inclusion between two subsets of U [9].

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

The function $\nu_{l,u}(X, Y) = f_{l,u}(\nu_{SRI}(X, Y))$, where

$$f_{l,u}(t) = \begin{cases} 0 & \text{if} \quad 0 \leq t \leq l \\ \frac{t-l}{u-l} & \text{if} \quad l < t < u \\ 1 & \text{if} \quad t \geq u \end{cases}$$

and $0 \leq l < u \leq 1$ is an example of a rough inclusion for the variable precision rough set model.

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

Definition 2 *For an approximation space* $AS_{\#,\$} = (U, I_{\#}, \nu_{\$})$ *and any subset* $X \subseteq U$ *the lower and the upper approximations are defined by*
$LOW\left(AS_{\#,\$}, X\right) = \{x \in U : \nu_{\$}(I_{\#}(x), X) = 1\}$,
$UPP\left(AS_{\#,\$}, X\right) = \{x \in U : \nu_{\$}(I_{\#}(x), X) > 0\}$, *respectively.*

Approximations of concepts (sets) are constructed on the basis of background knowledge. Obviously, concepts are also related to unseen so far objects. Hence it is very useful to define parameterized approximations with parameters tuned in the searching process for approximations of concepts. This idea is crucial for construction of concept approximations using rough set methods. In our notation $\#, \$$ are denoting vectors of parameters which can be tuned in the process of concept approximation.

3 Hierarchical System of Information Granules

In this section we construct a system of granules based on a given data table. We introduce syntax, semantics, inclusion and closeness of information granules.

We present hierarchical system of information granules based on the following four elements (see Figure 2):

- elementary granules corresponding to indiscernibility classes in the standard rough set model and to tolerance classes in the tolerance rough set model,
- decision rules,
- sets of decision rules,
- tolerance elementary granules.

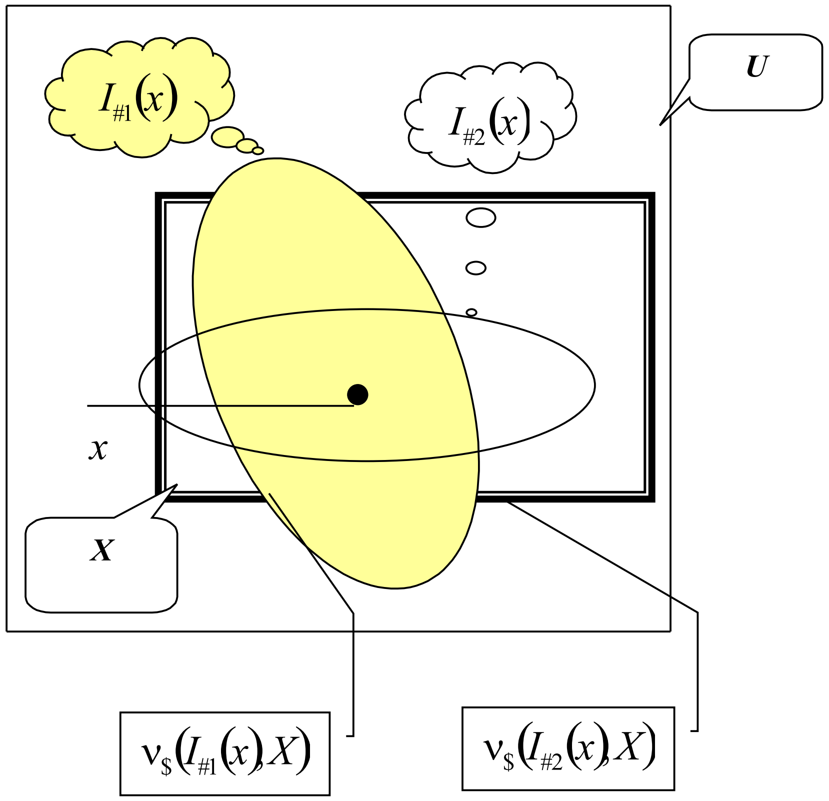

Figure 1: Elementary Granules

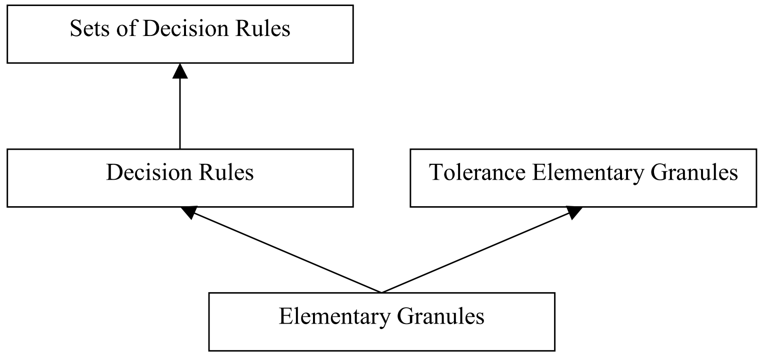

Figure 2: Hierarchy of Information Granules

3.1 Syntax and Semantics of Information Granules

In this subsection, we will consider several general kinds of information granules. We present now their syntax and semantics. In the following subsection we discuss the inclusion and closeness relations for information granules.

Syntax and semantics of elementary granules.

In an information system $IS = (U, A)$, elementary granules are defined by conjunctions of selectors (descriptors). For example $a \in V_1$ **and** $b \in V_2$, where $a, b \in A$.

For example, the meaning of an elementary granule $a \in V_1 \wedge b \in V_2$ is defined by

$$\|a \in V_1 \wedge b \in V_2\|_{IS} = \{x \in U : a(x) \in V_1 \; \& \; b(x) \in V_2\}.$$

Syntax and semantics of decision rules.

Let IS be an information system and let (α, β) be a new information granule received from the rule **if** α **then** β where α, β are elementary granules of IS. For example, one can consider rule of the form **if** $a \in V_1$ **and** $b \in V_2$ **then** $d = 1$.

The semantics $\|(\alpha, \beta)\|_{IS}$ of (α, β) is the pair of sets $(\|\alpha\|_{IS}, \|\beta\|_{IS})$. If the right hand sides of rules represent decision classes than among parameters to be tuned in classification is the number of conjuncts on the left hand sides of rules.

Syntax and semantics of sets of rules.

Assume that for some $k > 0$ there is a set of decision rules $Rule\_Set = \{(\alpha_i, \beta_i) : i = 1, \ldots, k\}$.

The semantics of $Rule\_Set$ is defined by

$$\|Rule\_Set\|_{IS} = \{\|(\alpha_i, \beta_i)\|_{IS} : i = 1, \ldots, k\}.$$

Syntax and semantics of tolerance elementary granules.

One can consider extension of elementary granules defined by tolerance relation. Let $IS = (U, A)$ be an information system and let τ be a tolerance relation (i.e. binary relation, which is reflexive and symmetric) on elementary granules of IS. Any pair (α, τ) is called a *τ-elementary granule.* The semantics $\|(\alpha, \tau)\|_{IS}$ of (α, τ) is the family $\{\|\beta\|_{IS} : (\beta, \alpha) \in \tau\}$.

3.2 Inclusion and Closeness of Information Granules

In this subsection we will discuss inclusion and closeness of different information granules introduced in the previous subsection. Let us mention that the choice of inclusion (closeness) definition depend very much on area of application and data analyzed. This is the reason that we have decided to introduce a separate subsection with this more subjective part of information granule semantics.

The inclusion relation between granules G, G' of degree at least p will be denoted by $\nu_p(G, G')$. Similarly, the closeness relation between granules G, G' of degree at least p will be denoted by $cl_p(G, G')$. By p we denote a vector of parameters (e.g. positive real numbers).

Inclusion and closeness of elementary granules.

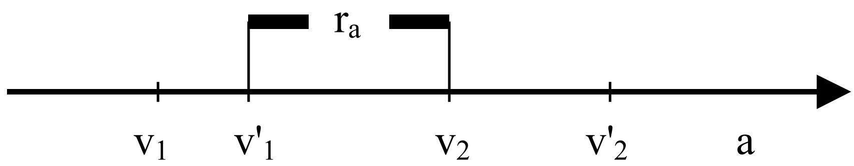

Figure 3: Overlapping Range r_a for Two Intervals $[v_1, v_2]$ and $[v'_1, v'_2]$

Let us consider for some real valued attribute a two elementary granules $a \in [v_1, v_2]$ and $a \in [v'_1, v'_2]$, where v_1, v_2, v'_1, v'_2 are real numbers and we assume that $v_1 < v_2$ and $v'_1 < v'_2$.

We define the overlapping range r_a (see Figure 3) for two intervals $[v_1, v_2]$ and $[v'_1, v'_2]$ by the following formula

$$r_a([v_1, v_2], [v'_1, v'_2]) = \max(\{\min(\{v_2, v'_2\}) - \max(\{v_1, v'_1\}), 0\}).$$

Elementary granule $a \in [v_1, v_2]$ is included in elementary granule $a \in [v'_1, v'_2]$ in degree t_a (in symbols, $\nu_{t_a}(a \in [v_1, v_2], a \in [v'_1, v'_2])$) if and only if

$$\frac{r_a([v_1, v_2], [v'_1, v'_2])}{v_2 - v_1} \geq t_a.$$

Two elementary granules $a \in [v_1, v_2]$ and $a \in [v'_1, v'_2]$ are close in degree t_a (in symbols, $cl_{t_a}(a \in [v_1, v_2], a \in [v'_1, v'_2])$) if and only if

$$\frac{1}{2}\left(\frac{r_a([v_1, v_2], [v'_1, v'_2])}{v_2 - v_1} + \frac{r_a([v_1, v_2], [v'_1, v'_2])}{v'_2 - v'_1}\right) \geq t_a.$$

Inclusion and closeness of decision rules.

Let us consider two decision rules $Rule$ and $Rule'$ with the same decision, for example:

$Rule$: **if** $a \in [v_1, v_2]$ **and** ... **then** $d = 1$,

$Rule'$: **if** $a \in [v'_1, v'_2]$ **and** ... **then** $d = 1$.

Let A_{Rule} and $A_{Rule'}$ be sets of attributes occurring in the **if** part of $Rule$ and $Rule'$, respectively.

$Rule$ is included in $Rule'$ in degree $(t, \{t_a : a \in A_{Rule} \cap A_{Rule'}\})$ if and only if

$$\frac{card(\{a \in A_{Rule} \cap A_{Rule'} : \nu_{t_a}(a \in [v_1, v_2], a \in [v'_1, v'_2])\})}{card(A_{Rule} \cup A_{Rule'})} \geq t.$$

Two rules $Rule$ and $Rule'$ are close in degree $(t, \{t_a : a \in A_{Rule} \cap A_{Rule'}\})$ if and only if

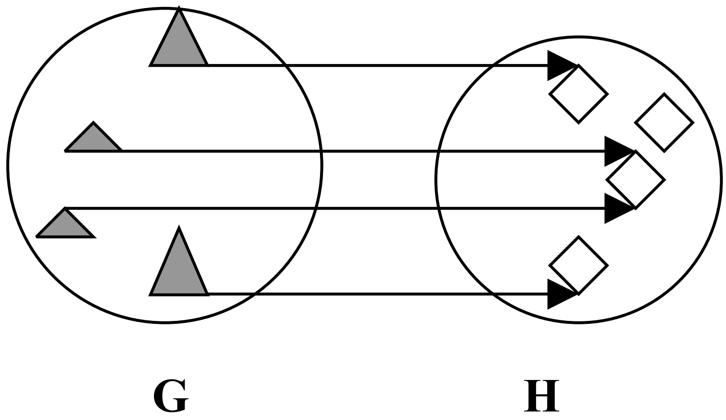

Figure 4: Inclusion for Sets G, H of Decision Rules

$$\frac{card(\{a \in A_{Rule} \cap A_{Rule'} : cl_{t_a}(a \in [v_1, v_2], a \in [v'_1, v'_2])\})}{card(A_{Rule} \cup A_{Rule'})} \geq t.$$

Inclusion and closeness of sets of rules.

The inclusion of $Rule\_Set$ in $Rule\_Set'$ can be defined by $\nu_{t,t'}^{IS}(Rule\_Set, Rule\_Set')$ if and only if for any $Rule \in Rule\_Set$ there is $Rule' \in Rule\_Set'$ for which $\nu_{t,t'}^{IS}(Rule, Rule')$

Granules $Rule\_Set$, $Rule\_Set'$ can be treated as close in a degree at least t (in IS) if and only if there exists a relation rel between sets of rules $Rule\_Set$ and $Rule\_Set'$ such that:

1. For any $Rule \in Rule\_Set$ there is $Rule' \in Rule\_Set'$ such that $(Rule, Rule') \in rel$ and $Rule$ is close to $Rule'$ (in IS) in degree at least t.

2. For any $Rule' \in Rule\_Set'$ there is $Rule \in Rule\_Set$ such that $(Rule, Rule') \in rel$ and $Rule$ is close to $Rule'$ (in IS) in degree at least t.

Inclusion for sets G, H of decision rules specified by inclusion of their elements is symbolized in Figure 4.

Inclusion and closeness of tolerance elementary granules.

For tolerance elementary granules, i.e., granules of the form (α, τ), (β, τ) one can consider the following inclusion measure:

$\nu_{t,t'}^{IS}((\alpha,\tau)(\beta,\tau))$ if and only if $\nu_{t,t'}^{IS}(\alpha',\beta')$ for any α',β' such that $(\alpha,\alpha')\in\tau$ and $(\beta,\beta')\in\tau$.

The closeness measure for tolerance elementary granules is defined by:

$cl_{t,t'}^{IS}((\alpha,\tau)(\beta,\tau))$ if and only if $\nu_{t,t'}^{IS}((\alpha,\tau)(\beta,\tau))$ and $\nu_{t,t'}^{IS}((\beta,\tau)(\alpha,\tau))$.

Conclusions

Our approach can be treated as a step towards understanding of complex information granules and their role in knowledge discovery. We have discussed information granule syntax and semantics as well their inclusion and closeness. Several examples of information granules have been presented. We have shown that some higher order patterns, important for knowledge discovery and data mining are expressible by means of complex information granules.

Acknowledgments

The research of Andrzej Skowron has been supported by the State Committee for Scientific Research of the Republic of Poland (KBN) research grant 8 T11C 025 19 and partially by the Wallenberg Foundation grant. The research of Jarosław Stepaniuk has been supported by the State Committee for Scientific Research of the Republic of Poland (KBN) research grant 8 T11C 025 19.

References

[1] Barsalou L.W.: Perceptual Symbol Systems, Behavioral and Brain Sciences Vol. 22, 1999, pp. 577–660.

[2] Escrig M.T., Toledo F.: Qualitative Spatial Reasoning: Theory and Practice, IOS Press, Amsterdam, 1998.

[3] Doherty P., Łukaszewicz W., Skowron A., Szałas A.: Knowledge Representation + Data Mining = Knowledge Mining (submitted).

[4] Düntsch I. (Ed.): Special issue of Fundamenta Informaticae on Spatial Reasoning (in press).

[5] Kargupta H., Chan Ph.: Advances in Distributed and Parallel Knowledge Discovery, AIII Press/MIT Press, Cambridge 2001.

[6] Pal S.K., Skowron A. (Eds.): Rough-Fuzzy Hybridization A New Trend in Decision Making, Springer–Verlag, 1999.

[7] Pawlak Z.: Rough Sets. Theoretical Aspects of Reasoning about Data, Kluwer Academic Publishers, Dordrecht, 1991.

[8] Polkowski L., Skowron A. (Eds.): Rough Sets in Knowledge Discovery 1 and 2. Physica-Verlag, Heidelberg, 1998.

[9] Skowron A., Stepaniuk J.: Tolerance Approximation Spaces, Fundamenta Informaticae, 27, 1996, pp. 245–253.

[10] Skowron A., Stepaniuk J.: Information Granules: Towards Foundations of Granular Computing, International Journal of Intelligent Systems, Vol. 16 No 1, 2001, pp. 57–86.

[11] Stepaniuk J.: Knowledge Discovery by Application of Rough Set Models, (Eds.) L. Polkowski, S. Tsumoto, T.Y. Lin, Rough Set Methods and Applications. New Developments in Knowledge Discovery in Information Systems, Physica–Verlag, Heidelberg, 2000, pp. 137–233.

[12] WITAS project web page: http://www.ida.liu.se/ext/witas/eng.html.

[13] Zadeh L.A., Kacprzyk J. (Eds.): Computing with Words in Information/Intelligent Systems, Vol.1–2, Physica–Verlag, Heidelberg, 1999.

[14] Zadeh L.A.: A New Direction in AI: Toward a Computational Theory of Perceptions, AI Magazine, Vol. 22 No 1, 2001, pp. 73–84.