1 Approximation Spaces in Rough Neurocomputing

Andrzej Skowron skowron@mimuw.edu.pl

Institute of Mathematics
Warsaw University
02-097, Banacha 2, Warsaw, Poland

Abstract. In the paper we discuss approximation spaces relevant for rough-neuro computing.

Keywords: rough sets, rough mereology, information granulation, information granules, rough neurocomputing

1.1 Introduction

Information sources provide us with granules of information that must be transformed, analyzed and built into structures that support problem solving. Lotfi A. Zadeh has recently pointed out to the need to develop a new research branch called Computing with Words (see, e.g., [14], [15], [16]). One way to achieve Computing with Words is through rough neurocomputing (see, e.g., [9], [3], [11], [4]) based on granular computing (GC) (see, e.g., [11]) and on rough neural networks performing computations on information granules rather than on real numbers. The main concepts of GC are related to information granule calculi. One of the main goals of information granule calculi is to develop algorithmic methods for construction of complex information granules from elementary ones by means of available operations and inclusion (closeness) measures. These constructions can also be interpreted as approximate schemes of reasoning (AR-schemes)(see, e.g., [11], [8]). Such schemes in distributed environments can be extended by adding interfaces created by approximation spaces. They make possible to induce approximations of concepts (or information about relations among them) exchanged between agents. In the paper we introduce approximation spaces as one of the basic concepts of rough neurocomputing paradigm.

A parameterized approximation space can be treated as an analogy to a neural network weight (see Fig. 1.1). In Fig. 1.1, $w_1, \cdots, w_n, \sum, f$ denote weights, aggregation operator, and activation function of a classical neuron, respectively, while $AS_1(P), \cdots, AS_n(P)$ denote parameterized approximations spaces where agents process input granules $G_1, \cdots, G_n$ and O denotes an operation (usually parameterized) that produces the output of a granular network. The parameters P of approximation spaces should be learned to induce the relevant information granules.

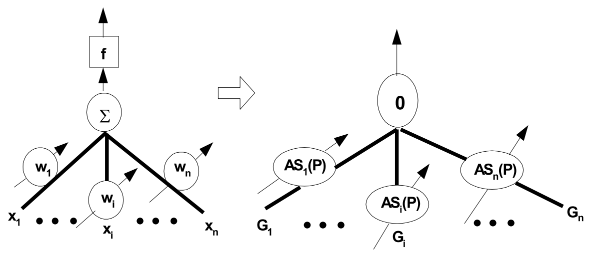

Fig. 1.1. Rough neuron

1.2 Approximation Spaces in Rough Set Theory

The starting point of rough set theory [5] is the indiscernibility relation, generated by information about objects of interest. The indiscernibility relation is intended to express the fact that due to the lack of knowledge we are unable to discern some objects employing the available information. It means that, in general, we are unable to deal with each particular object but we have to consider clusters of indiscernible objects, as fundamental concepts of our theory.

Suppose we are given two finite, non-empty sets U and A, where U is the *universe* of *objects, cases*, and A – a set of *attributes, features.* The pair $IS = (U, A)$ is called an *information table.* With every attribute $a \in A$ we associate a set V_a, of its *values*, called the *domain* of a. By $\mathbf{a}(x)$ we denote a data pattern $(a_1(x), \ldots, a_n(x))$ defined by the object x and attributes from $A = \{a_1, \ldots, a_n\}$. A data pattern of IS is any feature value vector $\mathbf{v} = (v_1, \ldots, v_n)$ where $v_i \in V_{a_i}$ for $i = 1, \ldots, n$ such that $\mathbf{v} = \mathbf{a}(x)$ for some $x \in U$.

Any subset B of A determines a binary relation $I(B)$ on U, called an *indiscernibility relation*, defined by $xI(B)y$ if and only if $a(x) = a(y)$ for every $a \in B$, where $a(x)$ denotes the value of attribute a for object x.

The family of all equivalence classes of $I(B)$, i.e., the partition determined by B, will be denoted by $U/I(B)$, or simply U/B; an equivalence class of $I(B)$, i.e., the block of the partition U/B, containing x will be denoted by $B(x)$.

If $(x, y) \in I(B)$ we will say that x and y are *B-indiscernible.* Equivalence classes of the relation $I(B)$ (or blocks of the partition U/B) are referred to as *B-elementary sets.* In the rough set approach the elementary sets are the basic building blocks (concepts) of our knowledge about reality. The unions of *B-elementary sets* are called *B-definable sets.*

The indiscernibility relation will be further used to define basic concepts of rough set theory. Let us define now the following two operations on sets

$$B_*(X) = \{x \in U : B(x) \subseteq X\},$$
$$B^*(X) = \{x \in U : B(x) \cap X \neq \emptyset\},$$

assigning to every subset X of the universe U two sets $B_*(X)$ and $B^*(X)$ called the *B-lower* and the *B-upper approximation* of X, respectively. The set

$$BN_B(X) = B^*(X) - B_*(X)$$

will be referred to as the *B-boundary region* of X.

If the boundary region of X is the empty set, i.e., $BN_B(X) = \emptyset$, then the set X is *crisp* (*exact*) with respect to B; in the opposite case, i.e., if $BN_B(X) \neq \emptyset$, the set X is referred to as *rough* (*inexact*) with respect to B.

1.3 Generalizations of Approximation Spaces

Several generalizations of rough set approach based on approximation spaces defined by (U, R), where R is an equivalence relation (called indiscernibility relation) in U, have been reported in the literature (for references see the papers and bibliography in ([7], [10], [12]).

Let us discuss in more details a definition of approximation space from [12].

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

- U is a non-empty set of objects,
- $I_{\#} : U \rightarrow P(U)$, where $P(U)$ denotes the powerset of U, is an uncertainty function,
- $\nu_{\$} : P(U) \times P(U) \rightarrow [0, 1]$ is a rough inclusion function.

If $p \in [0, 1]$ then $\nu_p(X, Y)$ denotes the following condition $\nu(X, Y) \geq p$ holds.

The uncertainty function defines for every object x a set of similarly described objects, i.e. the neighborhood $I_{\#}(x)$ of x. A constructive definition of uncertainty function can be based on the assumption that some metrics (distances) are given on attribute values.

A set $X \subseteq U$ is *definable in* $AS_{\#,\$}$, 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 [12]. For example, if X is non-empty then $\nu_p(X, Y)$ if and only if $p \leq \frac{card(X \cap Y)}{card(X)}$. If X is the empty set we assume $\nu_1(X, Y)$.

For a parameterized approximation space $AS_{\#,\$} = (U, I_{\#}, \nu_{\$})$ and any subset $X \subseteq U$ the lower and the upper approximations are defined by

$LOW\left(AS_{\#,\$}, X\right) = \left\{x \in U : \nu_{\$}\left(I_{\#}\left(x\right), X\right) = 1\right\}$,

$UPP\left(AS_{\#,\$}, X\right) = \{x \in U : \nu_{\$}\left(I_{\#}\left(x\right), X\right) > 0\}$, respectively.

In the case discussed i Section 1.2 above $I(x)$ is equal to the equivalence class $B(x)$ of the indiscernibility relation $I(B)$; in case when a tolerance (similarity) relation $\tau \subseteq U \times U$ is given we take $I(x) = \{y \in U : x\tau y\}$, i.e., $I(x)$ is equal to the tolerance class of τ defined by x. The standard inclusion relation is defined by $\nu(X, Y) = \frac{|X \cap Y|}{|X|}$ if X is non-empty, and otherwise $\nu(X, Y) = 1$. For applications it is important to have some constructive definitions of I and ν.

One can consider another way to define $I(x)$. Usually together with AS we consider some set F of formulae describing sets of objects in the universe U of AS defined by semantics $\| \cdot \|_{AS}$, i.e., $\|\alpha\|_{AS} \subseteq U$ for any $\alpha \in F$. Now, one can take the set

$$N_F(x) = \{\alpha \in F : x \in \|\alpha\|_{AS}\}$$

and $I(x) = \|\alpha\|_{AS}$ where α is selected or constructed from $N_F(x)$. Hence, more general uncertainty functions having values in $P(P(U))$ can be defined. The parametric approximation spaces are examples of such approximation spaces. These spaces have interesting applications. For example, by tuning of their parameters one can search for the optimal, under chosen criteria (e.g. the minimal description length), approximation space for concept description.

The approach based on inclusion functions has been generalized to the *rough mereological approach* [8]. The inclusion relation $x\mu_r y$ with the intended meaning *x is a part of y to a degree r* has been taken as the basic notion of the rough mereology being a generalization of the Leśniewski mereology. Rough mereology offers a methodology for synthesis and analysis of objects in distributed environment of intelligent agents, in particular, for synthesis of objects satisfying a given specification in satisfactory degree or for control in such complex environment. Moreover, rough mereology has been recently used for developing foundations of the *information granule calculus*, an attempt towards formalization of the Computing with Words paradigm, recently formulated by Lotfi Zadeh [14]. Research on rough mereology has shown importance of another notion, namely *closeness* of complex objects (e.g., concepts). This can be defined by $xcl_{r,r'}y$ if and only if $x\mu_r y$ and $y\mu_{r'}x$. The inclusion and closeness definitions of complex information granules are dependent on applications. However, it is possible to define the granule syntax and semantics as a basis for the inclusion and closeness definitions.

Finally let us mention that the approximation spaces are usually defined as parameterized approximation spaces. In the simplest case the parameter set is defined by the powerset of a given feature set. By parameter tuning the relevant approximation space is selected for given data set and target task.

1.4 Information Granule Systems and Approximation Spaces

In this section, we present a basic notion for our approach, i.e., information granule system. Any information granule system is any tuple

$$S = (G, R, Sem) \tag{1.1}$$

where

1. G is a finite set of parameterized constructs (e.g., formulas) called information granules;
2. R is a finite (parameterized) relational structure;
3. Sem is a semantics of G in R.

We assume that with any information granule system there are associated:

1. H a finite set of granule inclusion degrees with a partial order relation $<$ which defines on H a structure used to compare the inclusion degrees; we assume that H consists of the lowest degree 0 and the largest degree 1;
2. $\nu_p \subseteq G \times G$ a binary relation *to be a part to a degree at least* p between information granules from G, called *rough inclusion*. (Instead of $\nu_p(g, g')$ we also write $\nu(g, g') \geq p$.)

Components of an information granules system are parameterized. It means that we deal with parameterized formulas and a parameterized relational system. The parameters are tuned to make it possible to construct finally relevant information granules, i.e., granules satisfying specification or/and some optimization criteria. Parameterized formulas can consist of parameterized sub-formulas. The value set of parameters labeling a sub-formula is defining a set of formulas. By tuning parameters in optimization process or/and information granule construction a relevant subset of parameters is extracted and used for construction of the target information granule.

There are two kinds of computations on information granules. These are computations on information granule systems and computations on information granules in such systems, respectively. The first ones are aiming at construction of a relevant information granule systems defining parameterized approximation spaces for concept approximations used on different levels of target information granule constructions and the goal of the second ones is to construct information granules over such information granule systems to obtain target information granules, e.g., satisfying a given specification (at least to a satisfactory degree).

Examples of complex granules are tolerance granules created by means of similarity (tolerance) relation between elementary granules, decision rules, sets of decision rules, sets of decision rules with guards, information systems or decision tables (see, e.g., [8], [13], [11]). The most interesting class of information granules are information granules approximating concepts specified in natural language by means of experimental data tables and background knowledge.

One can consider as an example of the set H of granule inclusion degrees the set of binary sequences of a fixed length with the relation ν to be a part defined by the lexicographical order. This degree structure can be used to measure the inclusion degree between granule sequences or to measure the matching degree between granules representing classified objects and granules describing the left hand sides of decision rules in simple classifiers (see, e.g., [9]). However, one can consider more complex degree granules by taking as degree of inclusion of granule g_1 in granule g_2 the granule being a collection of common parts of these two granules g_1 and g_2.

New information granules can be defined by means of operations performed on already constructed information granules. Examples of such operations are set theoretical operations (defined by propositional connectives). However, there are other operations widely used in machine learning or pattern recognition ([2]) for construction of classifiers. These are the $Match$ and $Conflict\_res$ operations [9]. We will discuss such operations in the following section. It is worthwhile mentioning yet another important class of operations, namely, operations defined by data tables called decision tables [13]. From these decision tables, decision rules specifying operations can be induced. More complex operations on information granules are so called transducers [1]. They have been introduced to use background knowledge (not necessarily in the form of data tables) in construction of new granules. One can consider theories or their clusters as information granules. Reasoning schemes in natural language define the most important class of operations on information granules to be investigated. One of the basic problems for such operations and schemes of reasoning is how to approximate them by available information granules, e.g., constructed from sensor measurements.

In an information granule system, the relation ν_p to be a part to a degree at least p has a special role. It satisfies some additional natural axioms and additionally some axioms of mereology [6]. It can be shown that the rough mereological approach built on the basis of the relation to be a part to a degree generalizes the rough set and fuzzy set approaches. Moreover, such relations can be used to define other basic concepts like closeness of information granules, their semantics, indiscernibiliy and discernibility of objects, information granule approximation and approximation spaces, perception structure of information granules as well as the notion of ontology approximation. One can observe that the relation to be a part to a degree can be used to define operations on information granules corresponding to generalization of already defined information granules. For details the reader is referred to [4].

Let us finally note that new information granule systems can be defined using already constructed information granule systems. This leads to a hierarchy of information granule systems.

1.5 Classifiers as Information Granules

An important class of information granules create classifiers. One can observe that sets of decision rules generated from a given decision table $DT = (U, A, d)$ (see, e.g., [11]) can be interpreted as information granules. The classifier construction from DT can be described as follows:

1. First, one can construct granules G_j corresponding to each particular decision $j = 1, \ldots, r$ by taking a collection $\{g_{ij} : i = 1, \ldots, k_j\}$ of left hand sides of decision rules for a given decision.
2. Let E be a set of elementary granules (e.g., defined by conjunction of descriptors) over $IS = (U, A)$. We can now consider a granule denoted by
$$Match(e, G_1, \ldots, G_r)$$
for any $e \in E$ being a collection of coefficients ε_{ij} where $\varepsilon_{ij} = 1$ if the set of objects defined by e in IS is included in the meaning of g_{ij} in IS, i.e., $Sem_{IS}(e) \subseteq Sem_{IS}(g_{ij})$; and 0, otherwise. Hence, the coefficient ε_{ij} is equal to 1 if and only if the granule e matches in IS the granule g_{ij}.
3. Let us now denote by $Conflict\_res$ an operation (resolving conflict between decision rules recognizing elementary granules) defined on granules of the form $Match(e, G_1, \ldots, G_r)$ with values in the set of possible decisions $1, \ldots, r$. Hence, $Conflict\_res(Match(e, G_1, \ldots, G_r))$ is equal to the decision predicted by the classifier $Conflict\_res(Match(\bullet, G_1, \ldots, G_r))$ on the input granule e.

Hence, classifiers are special cases of information granules. Parameters to be tuned are voting strategies, matching strategies of objects against rules as well as other parameters like closeness of granules in the target granule.

The classifier construction is illustrated in Fig. 1.2 where three sets of decision rules are presented for the decision values $1, 2, 3$, respectively. Hence, we have $r = 3$. In figure to omit too many indices we write α_i instead of g_{i1}, β_i instead of g_{i2}, and γ_i instead of g_{i3}, respectively. Moreover, $\varepsilon_1, \varepsilon_2, \varepsilon_3$, denote $\varepsilon_{1,1}, \varepsilon_{2,1}, \varepsilon_{3,1}$; $\varepsilon_4, \varepsilon_5, \varepsilon_6, \varepsilon_7$ denote $\varepsilon_{1,2}, \varepsilon_{2,2}, \varepsilon_{3,2}, \varepsilon_{4,2}$; and $\varepsilon_8, \varepsilon_9$ denote $\varepsilon_{1,3}, \varepsilon_{2,3}$, respectively.

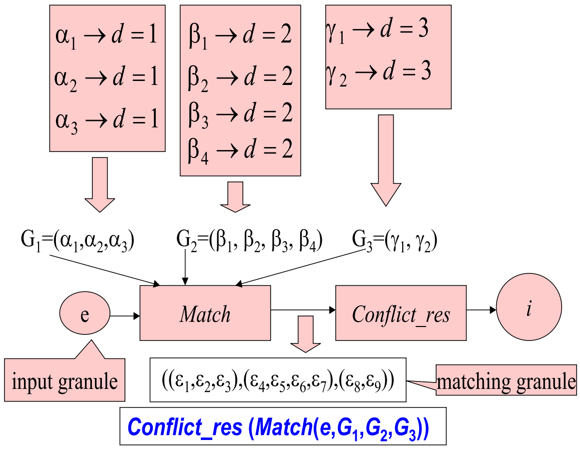

Fig. 1.2. Classifiers as Information Granules Granules

The reader can now easily describe more complex classifiers by means of information granules. For example, one can consider soft instead of crisp inclusion between elementary information granules representing classified objects and the left hand sides of decision rules or soft matching between recognized objects and left hand sides of decision rules.

1.6 Approximation Spaces for Information Granules

Using rough inclusions one can generalize the approximation operations for sets of objects, known in rough set theory, to arbitrary information granules.

The idea is to consider a family $G = \{g_t\}_t$ of granules by means of which a given granule g should be approximated. We assume that for a given set $\{g_1, \ldots, g_k\}$ of information granules included to a degree at least p in g there is a granule $Make\_granule(\{g_1, \ldots, g_k\})$ included to a degree at least $f(p)$ in g representing in a sense a collection $\{g_1, \ldots, g_k\}$ where f is a function transforming inclusion degrees into inclusion degrees. A typical example of $Make\_granule$ is set theoretical union used in rough set theory. We also assume inclusion degrees are partially ordered by a relation $<$.

Assume p is an inclusion degree, $G = \{g_t\}_t$ is a given family of information granules and g is a granule from a given information granules system S.

The (G, p)–*approximation* of g, in symbols $APP_{G,p}(g)$, is an information granule defined by

$$Make\_granule(\{g_t : \nu_p(g_t, g)\}).$$

Now, assuming $p < q$ one can consider two approximations for a given information granule g by G. The (G, q)–*lower approximation* of g is defined by

$LOW_{G,p}(g) = APP_{G,q}(g)$ (see Fig. 1.3). The (G, q)-*upper approximation* of g is defined by $UPP_{G,p}(g) = Make\_granule(\{APP_{G,p'}(g) : p' > p\})$.

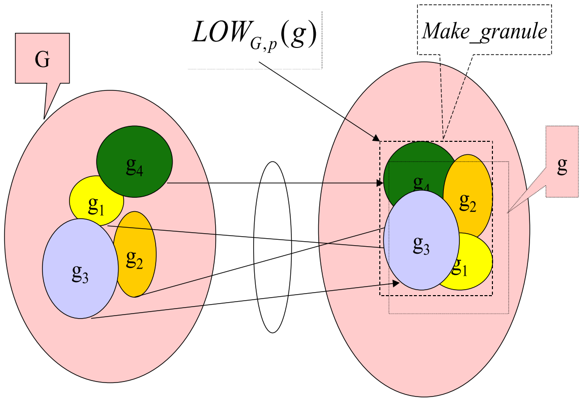

Fig. 1.3. (G, p)-Lower Approximation of Granule g

One can observe that the discussed in Section 1.3 definition of a parameterized approximation space is an example of the introduced notion of information granule approximation. It is enough to assume G to be the set of all neighborhoods $I_{\#}(x)$ for $x \in U$, $g \subseteq U$, $Make\_granule$ to be the set theoretical union, $p = 0$, $q = 1$.

It is useful to define parameterized approximations with parameters tuned in the searching process for approximations of concepts. This idea is crucial for methods of construction of concept approximations.

1.7 Approximation Spaces in Rough-Neuro Computing

In this section we would like to look more deeply on the structure of approximation spaces in the framework of information granule systems.

Such information granule systems are satisfying some conditions related to their information granules, relational structure as well as semantics. These conditions are the following ones:

1. Semantics consists of two parts, namely relational structure R and its extension R^*.
2. Different types of information granules can be identified: (i) object granules (denoted by x), (ii) neighborhood granules (denoted by n with subscripts), (iii) pattern granules (denoted by pat), and (iv) decision class granules (denoted by c).
3. There are decision class granules $c_1, \cdots, c_r$ with semantics in R^* defined by a partition of object granules into r decision classes. However, only the restrictions of these collections to the object granules from R are given.
4. For any object granule x there is a uniquely defined neighborhood granule n_x.

5. For any class granule c there is constructed a collection granule $\{(pat, p) : \nu_p^R(pat, c)\}$ of pattern granules labeled by maximal degrees to which pat is included in c (in R).
6. For any neighborhood granule n_x there is distinguished a collection granule $\{(pat, p) : \nu_p^R(n_x, pat)\}$ of pattern granules labeled by maximal degrees to which n_x is at least included in pat (in R).
7. There is a class of $Classifier$ functions transforming collection granules (corresponding to a given object x) described in two previous steps into the powerset of $\{1, \cdots, r\}$. One can assume object granules to be the only arguments of $Classifier$ functions if other arguments are fixed.

The classification problem is to find a $Classifier$ function defining a partition of object granules in R^* as close as possible to the partition defined by decision classes.

Any such $Classifier$ defines the lower and the upper approximations of union of decision classes c_i over $i \in I$ where I is a non-empty subset of $\{1, \cdots, r\}$ by

$$\underline{Classifier}(\{c_i\}_{i \in I}) = \{x \in \bigcup_{i \in I} c_i : \emptyset \neq Classifier(x) \subseteq I\}$$

$$\overline{Classifier}(\{c_i\}_{i \in I}) = \{x \in U^* : Classifier(x) \cap I \neq \emptyset\}.$$

The positve region of $Classifier$ is defined by

$$POS\,(Classifier) = \underline{Classifier}(\{c_1\}) \cup \cdots \cup \underline{Classifier}(\{c_r\}).$$

The closeness of the partition defined by the constructed $Classifier$ and the partition in R^* defined by decision classes can be measured, e.g., using ratio of the positive region size of $Classifier$ to the size of the object universe. The quality of $Classifier$ can be defined taking, as usual, only into account objects from $U^* - U$:

$$quality(Classifier) = \frac{card(POS\,(Classifier) \cap (U^* - U))}{card((U^* - U))}.$$

One can observe that approximation spaces have many parameters to be tuned to construct the approximation of class granules of high quality.

1.8 Conclusion

We have introduced the approximation space definition as one of the basic notions of rough neurocomputing paradigm. Approximation spaces can be treated as target information granule systems in which efficient search for relevant information granules (approximating concepts) can be performed. The approximation concept definition known from rough set theory [5] have been modified to capture inductive reasoning aspects in concept approximation.

Acknowledgements

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

References

1. P. Doherty, W. Łukaszewicz, A. Skowron, and A. Szałas. *Combining rough and crisp knowledge in deductive databases*. 2001 (submitted to [4]).
2. T.M. Mitchell. *Machine Learning*. Mc Graw-Hill, Portland, 1997.
3. S.K. Pal, W. Pedrycz, A. Skowron, and R. Swiniarski, editors. *Rough-Neuro Computing (special issue)*, volume 36. Elsevier, 2001.
4. S.K. Pal, L. Polkowski, and A. Skowron, editors. *Rough-Neuro Computing: Techniques for Computing with Words*. Springer-Verlag, Berlin, 2002. (in preparation).
5. Z. Pawlak. *Rough Sets. Theoretical Aspects of Reasoning about Data*. Kluwer Academic Publishers, Dordrecht, 1991.
6. L. Polkowski and A. Skowron. Rough mereology: a new paradigm for approximate reasoning. *International J. Approximate Reasoning*, 15/4:333–365, 1996.
7. L. Polkowski and A. Skowron, editors. *Rough Sets in Knowledge Discovery* Vol.1-2, Physica-Verlag, Heidelberg, 1998.
8. L. Polkowski and A. Skowron. Towards adaptive calculus of granules. In *[17]*, pages 201–227, 1999.
9. L. Polkowski and A. Skowron. Rough-neuro computing. *Lecture Notes in Artificial Intelligence*, pages 25–32, 2001.
10. A. Skowron. Rough sets in KDD. In Z. Shi, B. Faltings, and M. Musem, editors, *16-th World Computer Congress (IFIP'2000):Proceedings of Conference on Intelligent Information Processing (IIP2000)*, pages 1–17, Beijing, 2000. Publishing House of Electronic Industry. (plenary talk).
11. A. Skowron. Toward intelligent systems: Calculi of information granules. *Bulletin of the International Rough Set Society*, 5/1–2:9–30, 2001.
12. A. Skowron and J. Stepaniuk. Tolerance approximation spaces. *Fundamenta Informaticae*, 27:245–253, 1996.
13. A. Skowron and J. Stepaniuk. Information granules: Towards foundations of granular computing. *International Journal of Intelligent Systems*, 16/1:57–86, 2001.
14. L.A. Zadeh. Fuzzy logic = computing with words. *IEEE Trans. on Fuzzy Systems*, 4:103–111, 1996.
15. L.A. Zadeh. Toward a theory of fuzzy information granulation and its certainty in human reasoning and fuzzy logic. *Fuzzy Sets and Systems*, 90:111–127, 1997.
16. L.A. Zadeh. A new direction in ai: Toward a computational theory of perceptions. *AI Magazine*, 22/1:73–84, 2001.
17. L.A. Zadeh and J. Kacprzyk, editors. *Computing with Words in Information/Intelligent Systems*, volume 1-2. Physica-Verlag, Heidelberg, 1999.