Classifiers based on granular structures from rough inclusions

Lech Polkowski[1] **and Piotr Artiemjew**[2]
Polish–Japanese institute of Information Technology[1]
Warsaw, Poland, and University of Warmia and Mazury[2]
Olsztyn, Poland
polkow@pjwstk.edu.pl;artem@matman.uwm.edu.pl

Abstract

Granular computing initiated by L.A.Zadeh aims at computing with granules of knowledge i.e. with classes of objects similar with respect to a chosen representation of knowledge. It is assumed that classes of sufficiently similar objects behave satisfactorily similarly in solutions of problems of decision making, classification, fusion of knowledge, approximate reasoning. In this work, we support this assumption with a study of classifiers based on granular structures. Granules of knowledge are defined here by means of rough inclusions as proposed by Polkowski and from granules computed in this way in a data set, granular reflections are produced which are basis for inducing classifiers.

We demonstrate a few basic rough inclusions and we show that classifiers obtained from granulated according to them data sets yield results better than the standard exhaustive rough set classifier.

Keywords rough sets, rough inclusions, granules of knowledge, classification of data

1 Motivations:Rough set analysis of vagueness vs. mereology

Rough set analysis of vague concepts [6], begins with the idea of saturation of concepts by classes of indiscernibility: given an information function $Inf : U \to V$ defined on objects in a set U with values in a set V which induces an indiscernibility relation Ind on the set $U \times U$ with $Ind(u, v)$ if and only if $Inf(u) = Inf(v)$, concepts $X \subseteq U$ are divided into two categories: the category of *Inf–definable* concepts which are representable as unions of classes $[u]_{Ind} = \{v \in U : Ind(u, v)\}$ of the relation Ind, and the category of *Inf–non–definable* (or, *Inf–rough*) concepts which do not possess the definability property.

Definable concepts are the concepts which can be described with certainty: for each object $u \in U$, and a definable concept X, either u belongs in X or u does not belong in X; whereas for a non–definable concept Y, there exist objects u, v such that $Ind(u, v)$ and u belongs in Y but v belongs in $U \setminus Y$.

Rough set theory solves the problem of how to specify a non–definable concept with the idea of an approximation: given a concept Y, there exist by completeness of the containment relation $\subseteq$, two definable concepts $\underline{Y}$ and $\overline{Y}$ such that $\underline{Y} \subseteq Y \subseteq \overline{Y}$, $\underline{Y}$ is the largest definable subset of Y and $\overline{Y}$ is the smallest definable superset of Y.

The following points deserve attention in the above presented scheme:

1. Definable concepts are unions of atomic concepts: indiscernibility classes;

2. Non–definable concepts are approached with definable ones by means of containment.

Both operations involved in 1., 2., above, are particular cases of general constructs of mereology: the union of sets is a particular class operator and containment is a particular ingredient relation. It follows that setting the

L. Magdalena, M. Ojeda-Aciego, J.L. Verdegay (eds): Proceedings of IPMU'08, pp. 1786–1794
Torremolinos (Málaga), June 22–27, 2008

rough set context in the realm of mereology, one obtains a more general and formally adequate means of analysis.

The relation π of being a part is in mereology theory of Lesniewski [3] constructed as a non–reflexive and transitive relation on entities, i.e.,

p1. $\pi(u, u)$ for no entity u; p2. $\pi(u, v)$ and $\pi(v, w)$ imply $\pi(u, w)$.

An example is the proper containment relation $\subset$ on sets.

A standard usage of the Schroeder theorem makes π into a partial order relation ing of an *ingredient*: $v\ ing\ u$ if and only if either $\pi(v, u)$ or $v = u$. Clearly, ing is reflexive, weakly–antisymmetric and transitive. An example is the containment relation $\subseteq$ on sets.

The union of sets operator used in constructions of approximations, has its counterpart in the mereological class operator Cls [3]; it is applied to any non–empty collection F of entities to produce the entity $ClsF$; the formal definition is given in terms of the ingredient relation: an entity X is the class $ClsF$ if and only if the two conditions are satisfied,

c1. $u\ ing\ X$ for each $u \in F$;

c2. $u\ ing\ X$ implies the existence of entities v, w with the properties:

i. $v\ ing\ u$;

ii. $v\ ing\ w$;

iii. $w \in F$.

It is easy to verify that in case π is $\subset$, ing is $\subseteq$, F a non–empty collection of sets, the $ClsF$ is $\bigcup F$, the union of F. Class operator will be the principal tool in the definition of granules, see [10], [11], [12].

2 Rough mereology and rough inclusions: Motivation

In the process of development of rough set theory, it has turned out that indiscernibility should be relaxed to similarity: in [16] attention was focused on tolerance relations, i.e., relations which are reflexive and symmetric but need not be transitive. An example of such relation was given in [7]: given a metric ρ and a fixed small positive δ, one declares points x, y in the relation $sim(\delta)$ if and only if $\rho(x, y) < \delta$. The relation $sim(\delta)$ is a tolerance relation but it is equivalence for non–archimedean ρ's only.

We continue this example by introducing a graded version of $sim(\delta)$, viz., for a real number $r \in [0, 1]$, we define the relation $sim(\delta, r)$ by letting,

$$sim(\delta, r)(x, y) \text{ iff } \rho(x, y) \leq 1 - r. \quad (1)$$

The collection $sim(\delta, r)$ of relations have the following properties evident by the properties of the metric ρ:

sim1. $sim(\delta, 1)(x, y)$ iff $x = y$;

sim2. $sim(\delta, 1)(x, y)$ and $sim(\delta, r)(z, x)$ imply $sim(\delta, r)(z, y)$;

sim3. $sim(\delta, r)(x, y)$ and $s < r$ imply $sim(\delta, s)(x, y)$.

Properties sim1.–sim3. induced by the metric ρ refer to the ingredient relation $=$ whose corresponding relation of part is empty; a generalization can thus be obtained by replacing the identity with an ingredient relation ing in a mereological universe (U, π).

In consequence a relation $\mu(u, v, r)$ is defined that satisfies the following conditions:

rm1. $\mu(u, v, 1)$ iff $u\ ing\ v$;

rm2. $\mu(u, v, 1)$ and $\mu(w, u, r)$ imply $\mu(w, v, r)$;

rm3. $\mu(u, v, r)$ and $s < r$ imply $\mu(u, v, s)$.

Any relation μ which satisfies the conditions rm1.–rm3. is called a *rough inclusion*, see [9], [17]. This relation is a similarity relation which is not necessarily symmetric, but it is reflexive. It is read as "the relation of a part to a degree".

3 Granules of knowledge and granular reflections of data sets

The issue of granulation of knowledge as a problem on its own, has been posed by L.A.Zadeh, see [28], [29]. The issue of granulation has been a subject of intensive studies within rough set community, as witnessed by a number of papers, e.g., [4], [20], papers in [21].

Rough set context offers a natural venue for granulation, and indiscernibility classes were recognized as *elementary granules* whereas their unions serve as *granules of knowledge*; these granules and their direct generalizations to various similarity classes induced by general binary relations were subject to a research, see, e.g., [4], [27].

Granulation of knowledge by means of rough inclusions was studied in Polkowski and Skowron [17], [18], [19], [20] and in Polkowski [9], [10], [11].

The general scheme for inducing granules on the basis of a rough inclusion is as follows.

For an information system $I = (U, A)$ and a *rough inclusion* μ on U, for each object u and each real number $r \in [0, 1]$, we define the *granule* $g_\mu(u, r)$ *about* u *of the radius* r, *relative to* μ.

$$g_\mu(u, r) \text{ is } ClsF(u, r), \tag{2}$$

where the property $F(u, r)$ is satisfied with an object v if and only if $\mu(v, u, r)$ holds and Cls is the class operator of Mereology defined in (**??**).

It was shown in Polkowski (2004; Thm. 4), that in case of a transitive and symmetric μ,

$$v\ ing\ g_\mu(u, r) \text{iff } \mu(v, u, r). \tag{3}$$

Property (3) allows for representing the granule $g_\mu(u, r)$ as the list or a set of those objects v for which $\mu(v, u, r)$ holds.

For a given granulation radius r, and a rough inclusion μ, we form the collection $G_r^\mu = \{g_\mu(u, r)\}$ of all granules of the radius r relative to μ.

Granular data sets were proposed by L.Polkowski in [10],[11], [13] as the following constructions. Given $r \in [0, 1]$, the set of all granules $G_r^\mu = \{g_\mu(u, r) : u \in U\}$ is defined. From this set, a covering $Cov_r\mu h(\mathcal{G})$ is chosen according to a strategy $\mathcal{G}$. Granules in $Cov_r^\mu(\mathcal{G})$ form a new universe of objects. For each $g \in Cov_r^\mu(\mathcal{G})$, and each attribute $a \in A$, a factored attribute a_f is defined as $a_f(g) = \mathcal{S}(\{a(u) : u \in g\})$.

The new information system $I_r^\mu = (Cov_r^\mu(\mathcal{G}), \{a_f : a \in A\})$ is a *granular reflection* of the original information system I. The same procedure is applied to a decision system $D = (U, A, d)$ to form the reflection $D_r^\mu = (Cov_r^\mu(\mathcal{G}), \{a_f : a \in A\}, d_f)$. The object $o(g)$ defined for a granule g by means of $inf(o(g)) = \{(a_f, a_f(g)) : a \in A\}$ according to a strategy $\mathcal{S}$ is called an $\mathcal{S}$–*reflection of the granule* g; clearly, $o(g)$ need not be a real object in the training or test sets: clearly, one cannot prove the existence or non–existence of such an object in reality. Usage of these objects seems to be highly justified by their auxiliary role; they vanish in final effect playing the role of intermediary.

4 Classification of data by rough set tools

Given a data set (a decision system), $D = (U, A, d)$ where attributes in the set A induce on objects in the set U the information function $Inf_A(u) = \{(a, a(u)) : a \in A\}$ and the decision attribute d induced the information $Inf_d(u) = \{(d, d(u))\}$, a classifier is a set of judiciously chosen *decision rules* of the form $Inf_A(u) \Rightarrow Inf_d(u)$; we observe the duality between rules and objects: in both cases for each $a \in A$, the value $a(u)$, is defined. Thus each rule defines an object and each object defines a rule. We will avail of this duality in the sequel. Classification problem is to find a set if rules on basis of a given *training set* and to apply this set in finding decision classes for objects in a given *test set*.

In most general terms, building a classifier consists in searching in the pool of descriptors for their conjuncts that describe sufficiently well decision classes. As distinguished in [23], there are three main kinds of classifiers searched for: *minimal*, i.e., consisting of minimum possible number of rules describing decision classes in the universe, *exhaustive*, i.e., consisting of all possible rules, *satisfactory*, i.e., containing rules tailored to a specific use. In our exemplary classification task, the algorithm applied is the exhaustive algorithm supplied with the system RSES [22].

Classifiers are evaluated globally with respect to their ability to properly classify objects, usually by *error* which is the ratio of the number of correctly classified objects to the number of test objects, *total accuracy* being the ratio of the number of correctly classified cases to the number of recognized cases, and *total coverage*, i.e, the ratio of the number of recognized test cases to the number of test cases.

An important class of methods for classifier induction are those based on similarity or analogy reasoning; most generally, this method of reasoning assigns to an object u the value of an attribute a from the knowledge of values of a on a set $N(u)$ of objects whose elements are selected on the basis of a similarity relation, usually but not always based on an appropriate metric.

A study of algorithms based on similarity relations is [5]. The main tool in inducing similarity relations are *generalized templates*, i.e., propositional formulas built from *generalized descriptors* of the form $(a \in W_a)$ where W_a is a subset of the value set V_a as well as metrics like the Manhattan, Hamming, Euclidean, are used.

A realization of analogy–based reasoning idea is the *k–nearest neighbors* (k-nn) method in which for a fixed number k, values of a at k nearest to u objects in the training set. Finding nearest objects is based on some similarity measure among objects that in practice is a metric. Metrics to this end are built on the two basic metrics: the Manhattan metric for numerical values and the Hamming metric for nominal values, see, e.g., [25].

Our approach based on granulation can be also placed in similarity based reasoning as rough inclusions are measures of similarity degree among objects.

Table 1: Best results for Australian credit by some rough set based algorithms; in case $*$, reduction in object size is 40.6 percent, reduction in rule number is 43.6 percent; in case $**$, resp. 10.5, 5.9

so	*me*	*acc*	*cov*
[2]	$SNAPM(0.9)$	$error = 0.130$	–
[5]	$simple.templates$	0.929	0.623
[5]	$general.templates$	0.886	0.905
[5]	$closest.simple.templates$	0.821	.1.0
[5]	$closest.gen.templates$	0.855	1.0
[5]	$tolerance.simple.templ.$	0.842	1.0
[5])	$tolerance.gen.templ.$	0.875	1.0
[26]	$adaptive.classifier$	0.863	–
[14]	$granular^{*}.r = 0.642857$	0.867	1.0
[14]	$granular^{**}.r = 0.714826$	0.875	1.0
[1]	$conceptdependent.r = 0.785714$	0.9970	0.9995

In [14], results of experiments are reported with the data set *Credit card application approval data set (Australian credit)* [24] (UCI repository).

In rough set literature there are results of tests with other algorithms on Australian credit data set; we recall best of them in Table 1 and we give also best granular cases.

This Table does witness that granulation of knowledge applied in [14] as a new idea in data classification can lead to better results than the analysis based on individual objects. This confirms the validity of granular approach and reflects the fact that granularity is so important in Natural Reasoning.

Granules in this example have been computed with respect to the rough inclusion μ_h, computed in turn according to (1) from the reduced Hamming distance: $h(u,v) = \frac{|Inf_a(u) \cap Inf_A(v)|}{|A|}$; thus $\mu_(u,v,r)$ if and only if $|\{a \in A : a(u) = a(v)\}| \geq r \cdot |A|$ and the granule $g_r^h(v,r) = \{u \in U : \mu_h(u,v,r)$.

The last result in Table **??** refers to *concept dependent* granulation, see [1]: in this method granules are computed relative to decision classes, i.e., for each $v \in U$, $g_r^{h,cd}(v,r) = g_r^h(v,r) \cap \{u : d(u) = d(v)\}$.

5 Parameterized variants of rough inclusions μ_h in classification of data

For the formula $\mu_h(v,u,r)$ an extension is proposed which depends on a chosen metric ρ bounded by 1 in the attribute value space V

of (we assume for simplicity that ρ is suitable for all attributes).

Then, given an $\varepsilon \in [0,1]$, we let $\mu_h^{\varepsilon}(v,u,r)$ if and only if $|\{a \in A : \rho(a(v),a(u)) < \varepsilon\}| \geq r \cdot |A|$. It is manifest that μ^{ε} is a rough inclusion if ρ is a non–archimedean metric, i.e., $\rho(u,w) \leq max\{\rho(u,v),\rho(v,w)\}$; otherwise the monotonicity condition rm2 of sect. 2 need not be satisfied and this takes place with most popular metrics like Euclidean, Manhattan etc.

In this case, a rough inclusion is μ^* defined as follows: $\mu_h^*(v,u,r)$ if and only if there exists an ε such that $\mu_h^{\varepsilon}(v,u,r)$. Then it is easy to check that μ^* is a rough inclusion. The parameter r is called the catch radius.

Granules induced by the rough inclusion μ_h^* with $r=1$ have a simple structure: a granule $g_h^{\varepsilon}(u,1)$ consists of all $v \in U$ such that $\rho(a(u),a(v)) \leq \varepsilon$.

The idea poses itself to use granules defined in this way to assign a decision class to an object u in the test set. The implementation of this idea is as follows.

First on the training set, rules are induced by an exhaustive algorithm. Then, given a set *Rul* of these rules, and an object u in the test set, a granule $g_h^{\varepsilon}(u,1)$ is formed in the set *Rul*: in this, the duality between objects and rules is exploited as rules and objects can be written down in a same format of information sets. This also allows for using training objects instead of rules in forming granules and voting for decision by majority voting.

Thus, $g_h^{\varepsilon}(u,1) = \{r \in Rul : \rho(a(u),a(r)) \leq \varepsilon$for each attribute $a \in A$ where $a(r)$ is the value of the attribute a in the premise of the rule.

Rules in the granule $g_h^{\varepsilon}(u,1)$ are taking part in a voting process: for each value c of a decision class, the following factor is computed,

$$\text{param(c)} = \frac{\text{sum of supports of rules pointing to c}}{\text{cardinality of c in the training set}}. \quad (4)$$

cf. [2] for a discussion of various strategies of voting for decision values.

The class c_u assigned to u is decided by

$$param(c_u) = max_c param(c), \quad (5)$$

with random resolution of ties.

In computing granules, the parameter ε is normalized to the interval $[0,1]$ as follows: first, for each attribute $a \in A$, the value $train(a) = max_{training\ set}a - min_{training\ set}a$ is computed and the real line $(-\infty,+\infty)$ is contracted to the interval $[min_{training\ set}a, max_{training\ set}a]$ by the mapping f_a,

$$f_a(x) = \begin{cases} min_{training\ set}a \text{ in case } x \leq min_{training\ set}a \\ x \text{ in case } x \in [min_{training\ set}a, max_{training\ set}a] \\ max_{training\ set}a \text{ in case } x \geq max_{training\ set}a. \end{cases} \quad (6)$$

When the value $a(u)$ for a test object u is off the range $[min_{training\ set}a,\ max_{training\ set}a]$, it is replaced with the value $f_a(a(u))$ in the range. For an object v, or a rule r with the value $a(v)$, resp., $a(r)$ of a denoted $a(v,r)$, the parameter ε is computed as $\frac{|a(v,r)-f_a(a(u))|}{train(a)}$. The metric ρ was chosen as the metric $|x-y|$ in the real line. We show results of experiments with rough inclusions discussed in this work. Our data set was a subset of Australian credit data in which training set had 100 objects from class 1 and 150 objects from class 0 (which approximately yields the distribution of classes in the whole data set). The test set had 100 objects, 50 from each class. The RSES exhaustive classifier [22] applied to this data set gave accuracy of 0.79 and coverage of 1.0. In figures below this result of RSES is shown with a horizontal line at 0.79.

5.1 Results of tests with granules of training objects according to $\mu_h^{\varepsilon}(v,u,1)$ voting for decision

In Fig. 1 results of classification are given in function of ε for accuracy as well as for coverage.

5.2 Results of tests with granules of training objects according to $\mu_h^{\varepsilon}(v,u,r)$ voting for decision

We return to the rough inclusion $\mu_h^*(v,u,r)$ with general radius r. The procedure ap-

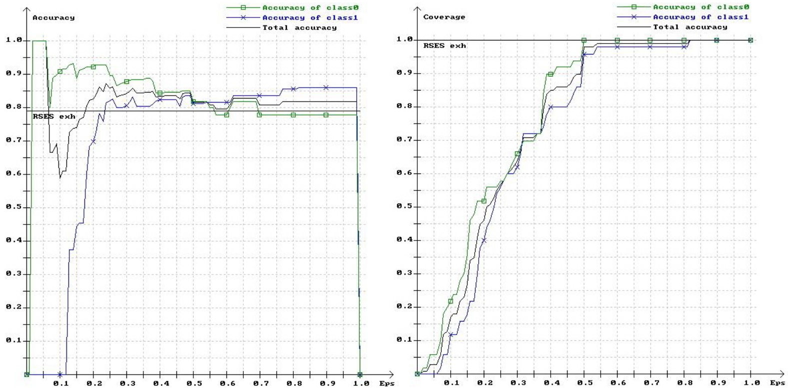

Figure 1: Results for algorithm 1_v1, Best result for $\varepsilon = 0.62$: accuracy = 0.828283, coverage = 0.99

Table 2: (40%-60%)(1-0); Australian credit; Algorithm 1_v2. r_catch=catch radius, optimal_eps=Best ε, acc= accuracy, cov= coverage

r_catch	*optimal eps*	*acc*	*cov*
nil	*nil*	0.79	1.0
0.071428	0	0.06	1.0
0.142857	0	0.66	1.0
0.214286	0.01	0.74	1.0
0.285714	0.02	0.83	1.0
0.357143	0.07	0.82	1.0
0.428571	0.05	0.82	1.0
0.500000	0	0.82	1.0
0.571429	0.08	0.84	1.0
0.642857	0.09	0.84	1.0
0.714286	0.16	0.85	1.0
0.785714	0.22	0.86	1.0
0.857143	0.39	0.84	1.0
0.928571	0.41	0.828283	0.99
1.000000	0.62	0.828283	0.99

plied in case of $\mu_h^{\varepsilon}(v, u, 1)$ can be repeated in the general setting. The resulting classifier is a function of two parameters ε, r. In Table 5.2 results are included where against values of the catch radius r the best value for ε's marked by the optimal value *optimal eps* is given for accuracy and coverage.

6 Rough inclusions and their weaker variants obtained from residual implications in classification of data

Residual implications of continuous t–norms can supply rough inclusions according to a general formula,

$$\mu_\phi(v, u, r) \text{ iff } \phi(u) \Rightarrow_t \phi(v) \geq r, \qquad (7)$$

where ϕ maps the set U of objects into $[0, 1]$ and $\phi(u) \leq \phi(v)$ if and only if $u\ ing\ v$ (ing is an ingredient relation of the underlying mereology, $\Rightarrow_t$ is the residual implication induced by the t–norm , i.e., $x \Rightarrow_t y \geq z$ if and only if $t(x, z) \leq y$, see [8].

A weak interesting variant of this class of rough inclusions is indicated. This variant uses sets $dis_\varepsilon(u, v) = \frac{|\{a \in A: \rho(a(u), a(v)) \geq \varepsilon\}|}{|A|}$, and $ind_\varepsilon(u, v) = \frac{|\{a \in A: \rho(a(u), a(v)) < \varepsilon\}|}{|A|}$,

for $u, v \in U$, $\varepsilon \in [0, 1]$, where ρ is a metric $|x - y|$ on attribute value sets.

The resulting weak variant of the rough inclusion μ_ϕ is,

$$\mu_t(u, v, r) \text{ iff } dis_\varepsilon(u, v) \rightarrow_t \ ind_\varepsilon(u, v) \geq r. \qquad (8)$$

Basic variants for three principal t–norms: the Łukasiewicz t–norm $L = max\{0, x + y - 1\}$, the product t–norm $P(x, y) = x \cdot y$, and $min\{x, y\}$ are, (the value in all variants is 1 if and only if $x \leq y$ so we give values only in the contrary case)

$$\mu_t(u, v, r) \text{ iff } \begin{cases} 1 - dis_\varepsilon(u, v) + ind_\varepsilon(u, v) \geq r \text{ for L} \\ \frac{ind_\varepsilon(u,v)}{dis_\varepsilon(u,v)} \geq r \text{ for P} \\ ind_\varepsilon(u, v) \geq r \text{ for min} \end{cases} \qquad (9)$$

Objects in the class c in the training set vote for decision at the test object u according to the formula: p(c)=$\frac{\sum_{v \in c} w(v,t)}{|c| \text{ in the training set}}$ where weight $w(v, t)$ is $dis_\varepsilon(u, v) \rightarrow_t ind_\varepsilon(u, v)$; rules induced from the training set pointing to the class c vote according to the

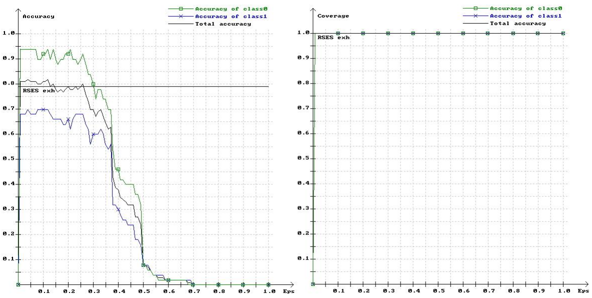

Figure 2: Results for algorithm 5_v1, Best result for $\varepsilon = 0.04$, accuracy = 0.82, coverage = 1

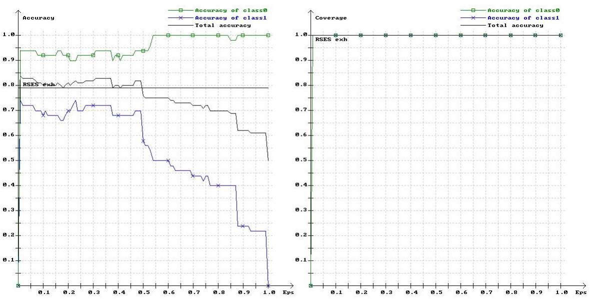

Figure 3: Results for algorithm 5_v2, Best result for $\varepsilon = 0.01$, accuracy = 0.84, coverage = 1

formula p(c)=$\frac{\sum_r w(r,t)\cdot support(r)}{|c| \text{ in the training set}}$. In either case, the class c* with p(c*)=max p(c) is chosen. We include here results of tests with training objects and t=min (Fig.2)and rules and t=min (Fig.3).

Similarly, we include in Figs. 4,5 results of tests with granules of training objects and rules for t=P, the product t–norm.

The results of tests in best cases for optimal values of ε exceed results obtained with the standard exhaustive algorithm.

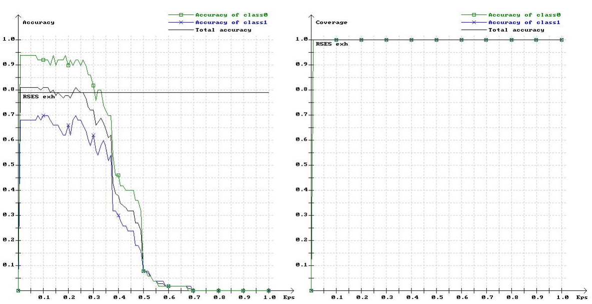

Figure 4: Results for algorithm 6_v1, Best result for $\varepsilon = 0.01$, accuracy = 0.81, coverage = 1

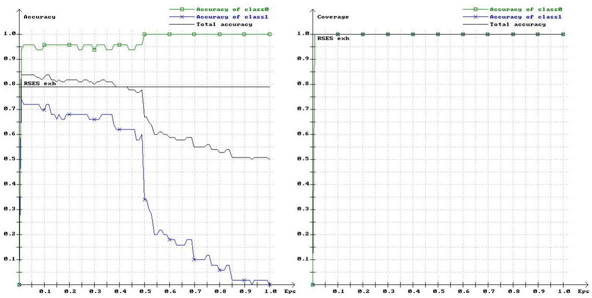

Figure 5: Results for algorithm 6_v2, Best result for *varepsilon* = 0.01, accuracy = 0.84, coverage = 1

7 Conclusion

The results shown here do witness that the proposed approach is a valid method for building effective classifiers and validates the hypothesis put forth in [10]. Highly dynamical procedure involving parameters r, ε reveals some tradeoffs between the size of a granule ε and the radius r of catching, whose optimal values depend as it may be conjectured on the structure of data. Further analysis will be devoted also to this aspect of our approach.

References

[1] P. Artiemjew, Classifiers from granulated data sets: Concept dependent and layered granulation, in: *Proceedings RSKD'07. Workshop at ECML/PKDD'07*, Warsaw Univ. Press, Warsaw, 2007, 1–9.

[2] J. G. Bazan, A comparison of dynamic and non–dynamic rough set methods for extracting laws from decision tables, In: *Rough Sets in Knowledge Discovery 1*, L. Polkowski, A.Skowron (Eds.), Physica Verlag, Heidelberg, 1998, 321–365.

[3] S. Leśniewski, On the foundations of set theory, *Topoi* 2, 1982, 7–52. A digest of S. Lesniewski, *On the Foundations of Set Theory*(in Polish), Moscow, 1916.

[4] Lin, T.Y., Granular computing: Examples,intuitions, and modeling, In: [21], 40–44.

[5] Nguyen Sinh Hoa, Regularity analysis and its applications in Data Mining,

In: *Rough Set Methods and Applications*, L.Polkowski, S.Tsumoto, T.Y.Lin (Eds.), Physica Verlag, Heidelberg, 2000, 289–378.

[6] Z. Pawlak, *Rough Sets: Theoretical Aspects of Reasoning about Data*, Kluwer, Dordrecht, 1991.

[7] H. Poincare, *Science et Hypothese*, Paris, 1905.

[8] L. Polkowski, *Rough Sets. Mathematical Foundations*, Physica Verlag, Heidelberg, 2002.

[9] L. Polkowski, Toward rough set foundations. Mereological approach (a plenary lecture), in: *Proceedings RSCTC04*, Uppsala, Sweden, 2004, LNAI vol. 3066,Springer Verlag, Berlin, 2004, 8–25.

[10] L. Polkowski, Formal granular calculi based on rough inclusions (a feature talk), In: [21], 57–62.

[11] L.Polkowski, Formal granular calculi based on rough inclusions (a feature talk), in: *Proceedings 2006 IEEE Intern. Conf. Granular Computing GrC'06*, Atlanta, USA, 2006, 57–62.

[12] L.Polkowski, The paradigm of granular rough computing, in: *Proceedings ICCI'07. 6th IEEE Intern. Conf. on Cognitive Informatics*, IEEE Computer Society, Los Alamitos, CA, 2007, 145–163.

[13] L. Polkowski, Granulation of knowledge in decision systems: The approach based on rough inclusions. The method and its applications (a plenary talk), In: *Proceedings RSEISP 07*, Warsaw, Poland, LNAI vol. 4585, Springer Verlag, Berlin, 2007.

[14] L. Polkowski, P.Artiemjew,, On granular rough computing: Factoring classifiers through granulated structures. In: *Proceedings RSEISP07*, LNAI vol. 4585, Springer Verlag, Berlin, 280–289.

[15] L. Polkowski, P. Artiemjew, On granular rough computing with missing values, In: *Proceedings RSEISP07*, LNAI vol. 4585, Springer Verlag, Berlin, 271–279.

[16] L. Polkowski, A.Skowron, J.Żytkow, Tolerance based rough sets, in: *Soft Computing:Rough Sets, Fuzzy Logic, Neural Networks, Uncertainty Management*, T.Y. Lin, M.A. Wildberger (Eds.), Simulation Councils, Inc., San Diego, 1995, 55–58.

[17] L.Polkowski, A. Skowron, Rough mereology: a new paradigm for approximate reasoning,*International Journal of Approximate Reasoning* 15(4), 1997, 333–365.

[18] L. Polkowski, A. Skowron, Grammar systems for distributed synthesis of approximate solutions extracted from experience, In: Gh. Paun and A.Salomaa (Eds.), *Grammatical Models of Multi–Agent Systems*, Amsterdam: Gordon and Breach, 316–333.

[19] L. Polkowski,A. Skowron, Towards an adaptive calculus of granules, In: L.A.Zadeh and J.Kacprzyk (Eds.), *Computing with Words in Information/Intelligent Systems,1.*, 201–228. Heidelberg: Physica Verlag.

[20] L. Polkowski, A. Skowron, Rough mereological calculi of granules: A rough set approach to computation, *Computational Intelligence. An International Journal*, 17(3), 2001, pp. 472–492.

[21] *Proceedings of IEEE 2005 Conference on Granular Computing GrC05*, Beijing, China, July 2005. IEEE Computer Society Press, 2005.

[22] A. Skowron et al., RSES; available at http://logic.mimuw.edu.pl/rses

[23] J. Stefanowski, On rough set based approaches to induction of decision rules. In L. Polkowski and Skowron, A. (Eds.), *Rough Sets in Knowledge Discovery 1*, Physica Verlag, Heidelberg, 1998, 500–529.

[24] http://www.ics.uci.edu./ mlearn/databases/

[25] A. Wojna, Analogy–based reasoning in classifier construction. *Transactions on Rough Sets IV*, subseries of Lecture Notes in Computer Science, vol. 3700, Springer Verlag, Berlin, 2005, 277–374.

[26] J. Wróblewski, Adaptive aspects of combining approximation spaces, In: S.K. Pal, L. Polkowski, A. Skowron (eds.),*Rough Neural Computing. Techniques for Computing with Words*, Springer Verlag, Berlin, 2004, 139–156.

[27] Yao, Y.Y., Perspectives of granular computing. In [21], 85–90.

[28] Zadeh, L.A.(1979). Fuzzy sets and information granularity. In M. Gupta, R. Ragade and R. Yager (Eds.) *Advances in Fuzzy Set Theory and Applications*, 3–18. Amsterdam: North–Holland.

[29] Zadeh, L.A. (2005). Graduation and granulation are keys to computation with information described in natural language. In [21], 30.