On Knowledge Granulation and Applications to Classifier Induction in the Framework of Rough Mereology

Lech Polkowski [1], Piotr Artiemjew [2]

[1] *Polish–Japanese Institute of Information Technology*
Koszykowa str. 86
Warszawa, 02–008, Poland *
E-mail: polkow@pjwstk.edu.pl

[2] *Department of Mathematics and Computer Science, University of Warmia and Mazury*
Zolnierska str. 14
Olsztyn, 10–560, Poland
E-mail: artem@matman.uwm.edu.pl

Received: 17/01/09
Accepted: 12/08/09

Abstract

Knowledge granulation as proposed by Zadeh consists in making objects under discussion into classes called granules; objects within a granule are similar one to another to a satisfactory degree relative to a chosen similarity measure. Rough mereology as developed by Polkowski in a series of works is especially suited to tasks of granulation as it does propose a systematic way to construct similarity measures in data sets and offers as well theoretic tools for granule formation in the form of an adaptation of the idea of mereological classes defined by Leśniewski in his mereology theory.

In this article, which extends our contributions to the Special Session on Rough Mereology organized by Polkowski and Artiemjew as a part of the Conference on Rough Sets and Knowledge Technology RSKT 2008, we give a fairly detailed account of basic ideas of rough mereology, a description of basic similarity measures called rough inclusions along with the idea of granulated data sets (granular reflections of data sets); then we follow with the idea on how to construct classifiers from granular data, and finally we present some results of granular classification on real data sets. In what follows, we restrict ourselves to a closed world of a given decision system, leaving aside metaphysical questions of relations between this system and the overwhelming universe of all feasible objects.

Keywords: rough sets, knowledge granulation, rough mereology, rough inclusions, classification of data into categories, granular data sets and classifiers

1. Rough set analysis of vagueness

It is well–known that rough set analysis of vague concepts (Refs. 1, 2), begins with the idea of saturation by classes of indiscernibility: given an information function $Inf : U \to V$ defined on objects in a finite set U with values in a set V which induces an indiscernibility relation *Ind* on the set $U \times U$ by re-

*This article is an extension of papers presented by Lech Polkowski and Piotr Artiemjew in the organized by them Special Session on Rough Mereology at the International Conference RSKT 2008 (Rough Sets and Knowledge Technology) at Chengdu, Sichuan, China quiring that $Ind(u,v)$ 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 above definability property.

Let us observe that definable concepts are those 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; it suffices to pick a $v \in X$, when X is non–empty, and for any given u to check whether $Inf(u) = Inf(v)$; 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*, cf. (Ref. 2): given a concept Y, there exist by completeness of the containment relation $\subseteq$, two definable concepts $\underline{Y}$ (the *lower approximation*) to Y, and $\overline{Y}$ (the *upper approximation*) to 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, cf. (Refs. 3, 4).

Remark 1.1. Definable concepts are unions of atomic concepts, viz., indiscernibility classes.

Remark 1.2. Non–definable concepts are approached with definable ones by means of set containment.

Both operations involved in Remarks 1.1, 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 regarding basics of rough set theory from mereology point of view, one obtains a more general and formally adequate means of analysis. This idea, pursued by Polkowski in a series of works (Refs. 3, 4, 5, 6, 7, 8, 9, 10, 11), is presented below in sect.4.

1.1. Data sets: Information/Decision Systems

Most often, indiscernibility relations are induced from the attribute–value description of objects in data by means of *information systems* (Refs. 1, 2); an information system is a a triple $I = (U, A, f)$ where U is a set of *objects*, A is a set of *attributes*, and f is a *value assignment* which for each attribute a and each object u assigns the value $f(a,u) \in V$ denoted shortly $a(u)$, where V denotes the *value set* in which attributes are evaluated. In this setting, the indiscernibility relation *Ind* is defined by means of information sets $Inf(u) = \{(a, a(u)) : a \in A\}$ as: $Ind(u,v)$ if and only if $Inf(u) = Inf(v)$, i.e., when $a(u) = a(v)$ for each $a \in A$.

Data sets are organized most often into *decision systems*, i.e., information systems in which an additional attribute, the *decision* d is added to *conditional attributes* in the set A: a decision system is then a quadruple $D = (U, A, d, f)$. The decision d is imposed by the real world (an expert) and attributes from A are meant to collectively yield as close to d approximation as possible. This is the subject of classification into *categories*: classes of $Ind(d)$.

To relate classes of $Ind(A)$ to classes of $Ind(d)$ in a *classifier*, the language of descriptors is in use (Refs. 2, 12). A *descriptor* is an elementary formula $(a = v)$, where $v \in a(U) \subseteq V$ is a value of a, which is interpreted in the set U as $[a = v] = \{u \in U : a(u) = v\}$.

Descriptors are combined into formulas of descriptor logic by means of sentential connectives $\vee$ (the disjunction), $\wedge$ (the conjunction), $\neg$ (the negation), $\Rightarrow$ (the implication), interpreted in the set U by means of recursive relations: $[\alpha \vee \beta] = [\alpha] \cup [\beta]$, $[\alpha \wedge \beta] = [\alpha] \cap [\beta]$, $[\neg\alpha] = U \setminus [\alpha]$; $[\alpha \Rightarrow \beta] = (U \setminus [\alpha]) \cup [\beta]$.

A *decision rule* is a formula,

$$r(B, d, \{v_a\}, v) : \bigwedge_{a \in B} (a = v_a) \Rightarrow (d = v), \quad (1)$$

where $B \subseteq A$ is a partial set of attributes.

Forming a decision rule means a search in the pool of available semantically non–vacuous descriptors for their combination that describes as closely as possible a chosen decision class.

Fulfilling this task, researchers on rough sets introduced some notions allowing for generation of rules with specific properties.

2. Classification

First, the idea of knowledge reduction has come, see (Ref. 2) for a discussion: a *reduct* B of the set A of attributes is a minimal subset of A with the property that $Ind(B) = Ind(A)$. An algorithm for finding reducts using methods of Boolean Reasoning was proposed by Skowron and Rauszer (Ref. 13); given input (U, A, f) with $U = \{u_1, ..., u_n\}$ it starts with the discernibility matrix,

$$M_{U,A} = [c_{i,j} = \{a \in A : a(u_i) \neq a(u_j)\}]_{1 \leqslant i,j \leqslant n}, \quad (2)$$

and builds the Boolean function,

$$g_{U,A} = \bigwedge_{c_{i,j} \neq \emptyset, i<j} \bigvee_{a \in c_{i,j}} \overline{a}, \quad (3)$$

where $\overline{a}$ is the Boolean variable assigned to the attribute $a \in A$.

The function $g_{U,A}$ is converted to its DNF form:

$$g^*_{U,A} : \bigvee_{j \in J} \bigwedge_{k \in K_j} \overline{a_{j,k}}. \quad (4)$$

Then: sets of the form $R_j = \{a_{j,k} : k \in K_j\}$ for $j \in J$, corresponding to prime implicants of $g^*_{U,A}$ are all reducts of A. Choosing a reduct R, and forming the reduced information system (U, R) one is assured that no information encoded in (U, A) has been lost.

Decision rules r are divided into *certain* (or, *exact*), when $[r] = U$, and possible, in the contrary case and when $[r] \neq \emptyset$; to induce the certain rule set, the notion of a *δ–reduct* was proposed by Skowron and Rauszer (Ref. 13); it is called a *relative reduct* in Bazan et al. (Ref. 14). To define δ–reducts, first the generalized decision δ_B is defined: for $u \in U$,

$$\delta_B(u) = \{v \in V_d : d(u') = v \wedge (u, u') \in ind(B) \wedge u' \in U\}. \quad (5)$$

A subset B of A is a δ–reduct to d when it is a minimal subset od A with respect to the property that $\delta_B = \delta_A$.

δ–reducts can be obtained from the modified Skowron and Rauszer algorithm (Ref. 13): it suffices to modify the entries $c_{i,j}$ to the discernibility matrix, by letting,

$$c^d_{i,j} = \{a \in A \cup \{d\} : a(u_i) \neq a(u_j)\}, \quad (6)$$

and then setting,

$$c'_{i,j} = \begin{cases} c^d_{i,j} \setminus \{d\} \text{ in case } d(u_i) \neq d(u_j) \\ \emptyset \text{ in case } d(u_i) = d(u_j). \end{cases} \quad (7)$$

The algorithm described above input with entries $c'_{i,j}$ forming the matrix $M^{\delta}_{U,A}$ outputs all δ–reducts to d encoded as prime implicants of the associated Boolean function $g^{\delta}_{U,A}$.

We write down a decision rule in the form $\phi/B, u \Rightarrow (d = v)$ where ϕ/B is a descriptor formula $\bigwedge_{a \in B}(a = a(u))$ over B. A method for inducing decision rules in a systematic way of Pawlak and Skowron (Ref. 15) and Skowron (Ref. 16) consists in finding the set of all δ–reducts $\mathbf{R}=\{R_1, ..., R_m\}$, and defining for each reduct $R_j \in \mathbf{R}$ and each object $u \in U$, the rule $\phi/R_j, u \Rightarrow (d = d(u)$. Rules obtained by this method are not minimal usually in the sense of the number of descriptors in the premise ϕ.

A method for obtaining *decision rules with minimal number of descriptors* (Ref. 16) consists in reducing a given rule $r : \phi/B, u \Rightarrow (d = v)$ by finding a set $R_r \subseteq B$ consisting of irreducible attributes in B only, in the sense that removing any $a \in R_r$ causes inequality $[\phi/R_r, u \Rightarrow (d = v)] \neq [\phi/R_r \setminus \{a\}, u \Rightarrow (d = v)]$ to hold. In case $B = A$, reduced rules $\phi/R_r, u \Rightarrow (d = v)$ are called *optimal basic rules* (with minimal number of descriptors). The method for finding of all irreducible subsets of the set A (Ref. 16) consists in considering another modification of discernibility matrix: for each object $u_k \in U$, the entry $c'_{i,j}$ into the matrix $M^{\delta}_{U,A}$ for δ–reducts is modified into,

$$c^k_{i,j} = \begin{cases} c'_{i,j} \text{ in case } d(u_i) \neq d(u_j) \text{ and } i = k \vee j = k \\ \emptyset \text{ otherwise.} \end{cases} \quad (8)$$

Matrices $M^k_{U,A}$ of entries $c^k_{i,j}$ and associated Boolean functions $g^k_{U,A}$ for all $u_k \in U$ allow for finding all irreducible subsets of the set A and in consequence all basic optimal rules (with minimal number of descriptors).

Decision rules are induced from a part of the decision system called the *training set* and they are judged by their quality in classifying new unseen as yet objects, i.e., by their performance on the remaining part of the decision system – the *test set* (Ref. 17). Quality evaluation is done on the basis of some measures: for a rule $r : \phi \Rightarrow (d = v)$, and an object $u \in U$, one says that u *matches* r in case $u \in [\phi]$. $match(r)$ is the number of objects matching r. *Support* $supp(r)$ of r is the number of objects in $[\phi] \cap [(d = v)]$; the fraction $cons(r)$=$\frac{supp(r)}{match(r)}$ is the *consistency degree* of r: $cons(r) = 1$ means that the rule is certain.

Strength, $strength(r)$, of the rule r is defined, see Michalski et al. (Ref. 18), Bazan (Ref. 19), Grzymala–Busse and Ming Hu (Ref. 20), as the number of objects correctly classified by the rule in the training phase; *relative strength* is defined as the fraction $rel - strength(r)$= $\frac{supp(r)}{|[(d=v)]|}$. *Specificity* of the rule r, $spec(r)$, is the number of descriptors in the premise ϕ of the rule r (Ref. 19).

In the testing phase, rules vie among themselves for object classification when they point to distinct decision classes; in such case, negotiations among rules or their sets are necessary. In these negotiations rules with better characteristics are privileged.

For a given decision class $c : [d = v]$, and an object u in the test set, the set $Rule(c,u)$ of all rules matched by u and pointing to the decision v, is characterized globally by $Support(Rule(c,u))$=$\sum_{r \in Rule(c,u)} strength(r) \cdot spec(r)$. The class c for which $Support(Rule(c,u))$ is the largest wins the competition and the object u is classified into the class $c : d = v$, see (Ref. 20).

It may happen that no rule in the available set of rules is matched by the test object u and partial matching is necessary, i.e., for a rule r, the *matching factor* $match - fact(r,u)$ is defined as the fraction of descriptors in the premise ϕ of r matched by u to the number $spec(r)$ of descriptors in ϕ. The rule for which the partial support $Part - Support(Rule(c,u))$=$\sum_{r \in Rule(c,u)} match - fact(r,u) \cdot strength(r) \cdot spec(r)$ is the largest wins the competition and it does assign the value of decision to u, see Grzymala–Busse and Ming Hu (Ref. 20).

In a similar way, notions based on relative strength can be defined for sets of rules and applied in negotiations among them (as discussed in Bazan et al. (Ref. 14).

As distinguished in Stefanowski (Ref. 21), there are three main kinds of classifiers searched for: *minimal*, i.e., consisting of minimum possible number of rules describing decision classes, *exhaustive*, i.e., consisting of all possible rules, *satisfactory*, i.e., containing rules tailored to a specific use. 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.

Minimum size algorithms include LEM2 algorithm due to Grzymala–Busse (Ref. 22) and covering algorithm in RSES package (Ref. 23); exhaustive algorithms include, e.g., LERS system due to Grzymala–Busse (Ref. 22) and systems based on discernibility matrices and Boolean reasoning by Skowron (Ref. 16), Bazan (Ref. 19), Bazan et al. (Ref. 14), implemented in the RSES package (Ref. 23). See also Stefanowski (Ref. 24) for a discussion of recent advances.

3. Similarity

Analysis based on indiscernibility has allowed for extracting the most important notions of rough set theory; further progress has been obtained by departing from indiscernibility to more general similarity relations. There have been various methods for introducing similarity relations.

An attempt at introducing some degrees of comparison among objects with respect to particular concepts consisted in defining *rough membership functions* in Pawlak and Skowron (Ref. 25): for an object u, an attribute set B and a concept $X \subseteq U$, the value,

$$\mu_{B,X}(u) = \frac{|[u]_B \cap X|}{|[u]_B|}, \tag{9}$$

was defined.

Informally, one can say that objects u, v are ε, B–similar with respect to the concept X in case $|\mu_{B,X}(u) - \mu_{B,X}(v)| < \varepsilon$. This relation is reflexive and symmetric, i.e., it is a *tolerance relation* for each ε, B, X, see Poincaré (Ref. 26) and Zeeman (Ref. 27).

Tolerance relations were introduced into rough sets in Nieminen (Ref. 28), and also studied in Polkowski, Skowron, and Żytkow (Ref. 29), Słowiński and Vanderpooten (Ref. 30), among others. It is possible to build a parallel theory of rough sets based on tolerance or similarity relations in analogy to indiscernibility relations. $\mu_{B,X}$ does characterize partial containment of objects in U into concepts X; a further step consists in considering general relations of partial containment in the form of predicates "*to be a part of to a degree at least*".

A general form of partial containment was proposed as an extension of mereological theory of concepts due to Lesniewski (Ref. 31); mereology takes the predicate *"to be a part of"* as its primitive notion, requiring of it to be irreflexive and transitive on the objects in the set U. The primitive notion of a *(proper) part* is relaxed to the notion of an *ingredient (an improper part) ing* $=$ *part* $\cup$ "$=''$.

The extension consists in considering the predicate *"to be a part to a degree of a least"*, formally introduced as the generic notion of a *rough inclusion* μ in Polkowski and Skowron (Ref. 32), see also (Refs. 33, 34, 35), as a ternary predicate (relation) with the semantic domain of $U \times U \times [0,1]$, discussed in detail below, in sect. 5.

Other methods for introducing similarity relations into realm of information systems, include methods based on *templates* and on *quasi–distance functions* in Nguyen S. H. (Ref. 36); a template is any conjunct of the form $T : \bigwedge_{a \in B}(a \in W_a)$ where B is a set of attributes, and $W_a \subseteq V_a$ is a subset of the value set V_a of the attribute a. Semantics of templates is defined as with descriptor formulas in sect.1. Templates are judged by some parameters: *length*, i.e., the number of generalized descriptors $(a \in W_a)$; *support*, i.e, number of matching objects; *approximated length (applength)*, i.e, $\sum_{a \in B} \frac{1}{|W_a \cap a(U)|}$. Quality of the template is given by a combination of some of parameters, e.g., *quality(T)=support(T)+applength(T)*. Templates are used in classification problems in the way analogical to decision rules.

A quasi–metric (a similarity measure) (Ref. 36) is a family $\Delta : \{\Delta_a : a \in A\}$ of functions where $\Delta_a(u,u) = 0$ and $\Delta_a(u,v) = \Delta_a(v,u)$ for $u, v \in U$. By means of these functions tolerance relations are built with help of standard metric–forming operators like max, $\sum$: $\tau_1(u,v) \Leftrightarrow max_a\{\Delta_a(u,v) \leqslant \varepsilon\}$, $\tau_2(u,v) \Leftrightarrow \sum_a \Delta_a(u,v) \leqslant \varepsilon\}$ for a given threshold ε are examples of such similarity relations. These similarity relations are applied in (Ref. 36) towards classifier construction.

4. Mereological analysis of vagueness

The fundamental relation π of *being a part* is in mereology theory of Lesniewski (Ref. 31) constructed as a non–reflexive and transitive relation on entity set U, i.e.,

Part 1. $\pi(u,u)$ for no entity u.

Part 2. $\pi(u,v)$ and $\pi(v,w)$ imply $\pi(u,w)$.

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

One makes π into a partial order relation *ing* of an *ingredient* by letting,

ing(u,v) if and only if either $\pi(u,v)$ or $u = v$.

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* (Ref. 36); 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,

Class 1. u *ing* X for each $u \in F$.

Class 2. 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 when $\pi = \subset$, hence *ing* = $\subseteq$, and F is a non–empty collection of sets, $ClsF$ is $\bigcup F$, the union of sets in the collection F.

Mereological reasoning about concepts and classes rests, to a substantial degree, on the Leśniewski Inference Rule (IR) (Ref. 31).

(IR) For entities x, y, if for each entity z, from z *ing* x it follows that there exists an entity w such that w *ing* z, w *ing* y, then x *ing* y.

Operators of the form *ClsF* are instrumental in our definition of granules of knowledge, discussed below in sect. 6.

4.1. A mereological extension of rough set theory

It follows from the previous sections, notably Remarks 1.1, 1.2, sect.1, that a more general rough set theory can be formulated by using mereological constructs in place of set theoretic ones. Given a collection $\mathscr{F}$ of concepts, and a part relation π along with the ingredient relation *ing*, we define an $(\mathscr{F},\pi)$*–definable* concept X as the class $Cls\ \mathscr{G}$ for some non–empty $\mathscr{G} \subseteq \mathscr{F}$.

Given a concept $Y \subseteq U$, the $(\mathscr{F},\pi)$*–lower approximation* $\underline{Y}(\mathscr{F},\pi)$ is defined as the class $Cls\ L(Y,\mathscr{F},\pi)$ where the property $L(Y,\mathscr{F},\pi)$ is satisfied by a concept Z in case Z *ing* Y. The upper approximation is defined via the complement and the lower approximation in the standard way, i.e., $\overline{Y}(\mathscr{F},\pi) = ClsF \setminus L(ClsF\setminus,\mathscr{F},\pi)$.

5. Rough mereology

In the process of development of rough set theory, it has been understood that indiscernibility relations could be replaced with more general and flexible similarity relations.

An inspiring example of such relation was given in (Ref. 26): given a metric ρ on a set U and a fixed small positive δ, one declares points x, y to be in the relation $sim(\delta)$ if and only if $\rho(x,y) < \delta$:

$$sim(\delta)(x,y) \text{ if and only if } \rho(x,y) < \delta. \qquad (10)$$

The relation $sim(\delta)$ is a tolerance relation.

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,

$$sim(\delta,r)(x,y) \text{ if and only if } \rho(x,y) \leqslant 1-r. \qquad (11)$$

The collections $sim(\delta,r)$ for $r \in [0,1]$, of relations, have the following properties evident by properties of the metric ρ.

Sim 1. $sim(\delta,1)(x,y)$ if and only if $x = y$.

Sim 2. $sim(\delta,1)(x,y)$ and $sim(\delta,r)(z,x)$ imply $sim(\delta,r)(z,y)$.

Sim 3. $sim(\delta,r)(x,y)$ and $s < r$ imply $sim(\delta,s)(x,y)$.

Properties Sim 1 – Sim 3, 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:

RM 1. $\mu(u,v,1)$ if and only if $ing(u,v)$.

RM 2. $\mu(u,v,1)$ and $\mu(w,u,r)$ imply $\mu(w,v,r)$.

RM 3. $\mu(u,v,r)$ and $s < r$ imply $\mu(u,v,s)$.

Any relation μ which satisfies the conditions RM 1 – RM 3 is called a *rough inclusion* (Refs. 4, 32). This relation is a similarity relation which is reflexive but not necessarily symmetric or transitive. It is read as "*the relation of a part to a degree at least of r*".

5.1. Rough inclusions: Case of information systems

The problem of methods by which rough inclusions could be introduced in information/decision systems has been studied, e.g., in (Refs. 4, 5, 6, 7, 8). Here we recapitulate these results for the convenience of the reader. We recall that an *information system* is a method of representing knowledge about a certain phenomenon in the form of a table of data; formally, it is a triple (U,A,f) as indicated in sect. 1.

5.1.1. Rough inclusions from metrics

As equation (11) shows, any metric ρ defines a rough inclusion μ_ρ by means of the equivalence $\mu_\rho(u,v,r) \Leftrightarrow \rho(u,v) \leqslant 1-r$. A very important example of a rough inclusion obtained on these lines see, e.g., (Refs. 3, 4, 5, 6, 7, 8) is the rough inclusion μ_h with $h(u,v)$ being the *reduced Hamming distance on information vectors* of u and v, i.e.,

$$h(u,v) = \frac{|\{a \in A : (a,a(u)) \neq (a,a(v))\}|}{|A|}, \qquad (12)$$

where the symbol $|A|$ denotes cardinality of the set A.

Thus, $\mu_h(u,v,r)$ if and only if $h(u,v) \leqslant 1-r$; introducing as in (Refs. 3–8) sets $DIS(u,v) = \{a \in A : (a,a(u)) \neq (a,a(v))\}$ and $IND(u,v) = A \setminus DIS(u,v)$=$\{a \in A : a(u) = a(v)\}$, along with quotients,

$$dis(u,v) = \frac{|DIS(u,v)|}{|A|}, \tag{13}$$

and

$$ind(u,v) = \frac{|IND(u,v)|}{|A|}, \tag{14}$$

one can write down the formula for μ_h either as,

$$\mu_h(u,v,r) \Leftrightarrow dis(u,v) \leqslant 1-r, \tag{15}$$

or,

$$\mu_h(u,v,r) \Leftrightarrow ind(u,v) \geqslant r. \tag{16}$$

Formula (16) witnesses that the rough inclusion μ_h is an extension of the indiscernibility relation *Ind* to a graded indiscernibility μ_h.

In computing μ_h, one can use in place of the metric h, any other metric function built on descriptors.

Rough inclusions induced by metrics possess an important property of *functional transitivity* (Ref. 4) expressed in a general form by the rule,

$$\frac{\mu_\rho(u,v,r), \mu_\rho(v,w,s)}{\mu_\rho(u,w,L(r,s))}, \tag{17}$$

where $L(r,s) = max\{0, r+s-1\}$ is the Łukasiewicz t–norm, see, e.g. (Ref. 37). We recall a short proof of this fact from (Ref. 4): assume that $\mu_\rho(u,v,r), \mu_\rho(v,w,s)$ which means in terms of the metric ρ that $\rho(u,v) \leqslant 1-r, \rho(v,w) \leqslant 1-s$; by the triangle inequality, $\rho(u,w) \leqslant (1-r)+(1-s)$, i.e., $\mu_\rho(u,w,r+s-1)$.

5.1.2. *Rough inclusions from functors of many–valued logics: the case of non–archimedean rough inclusions*

A function $t : [0,1] \times [0,1] \to [0,1]$ is a *t–norm*, see, e.g., (Refs. 37, 38), when it is increasing in each coordinate, symmetric (meaning $t(x,y) = t(y,x)$), associative (meaning $t(x,t(y,z)) = t(t(x,y),z)$, and it satisfies boundary conditions $t(x,0) = 0, t(x,1) = x$; a t–norm is *non–archimedean* in case the equality $t(x,x) = x$ holds for $x = 0,1$ only; it is known, see, e.g. (Refs. 3, 38), that only such t–norms up to an automorphism of the interval $[0,1]$ are the Łukasiewicz L and the product t–norm $P(x,y) = x \cdot y$.

Non–archimedean t–norms admit a functional representation: $t(x,y) = g(f(x)+f(y))$ (Ref. 39), where $f : [0,1] \to [0,1]$ is a continuous decreasing function, and g is the pseudo–inverse (Ref. 39) to f.

In the context of non–archimedean t–norms, one defines a rough inclusion μ_t by letting (Refs. 4–8),

$$\mu_t(u,v,r) \Leftrightarrow g(dis(u,v)) \geqslant r. \tag{18}$$

In particular, in case of the t–norm L, one has $g(x) = 1-x$, see (Ref. 39) , and thus the rough inclusion μ_L is expressed by means of the formula (16), i.e., it does coincide with the rough inclusion μ_h.

The scarcity of the stock of non–archimedean inclusions makes effectively only μ_L available as a granulation tool. In order to enlarge this collection, we reach to residua of continuous t–norms as first indicated in (Refs. 5, 6).

5.1.3. *Rough inclusions from functors of many–valued logic: the case of residual implications*

Other systematic method for defining rough inclusions is by means of residual implications of continuous t–norms (Refs. 5, 6).

For a continuous t–norm t, the *residual implication* $x \Rightarrow_t y$ is a mapping from the square $[0,1]^2$ into $[0,1]$ defined as follows, see, e.g., (Refs. 37, 38),

$$x \Rightarrow_t y \geqslant z \text{ if and only if } t(x,z) \leqslant y. \tag{19}$$

Thus, $x \Rightarrow_t y = max\{z : t(x,z) \leqslant y\}$.

Proposition 1. . *For every continuous t–norm t, the residual implication $x \Rightarrow_t y$ does induce a rough inclusion $\mu_t^{\Rightarrow}$ by means of the formula: $\mu_t^{\Rightarrow}(x,y,r)$ if and only if $x \Rightarrow_t y \geqslant r$.* **Proof.** We include a short argument for the sake of completeness; clearly, $\mu_t^{\Rightarrow}(x,x,1)$ holds as $x \Rightarrow_t y \geqslant 1$ is equivalent to $x \leqslant y$. Assuming $\mu_t^{\Rightarrow}(x,y,1)$, i.e., $x \leqslant y$, and $\mu_t^{\Rightarrow}(z,x,r)$, i.e., $z \Rightarrow x \geqslant r$ hence $t(z,r) \leqslant x$ we have $t(z,r) \leqslant y$, i.e., $z \Rightarrow y \geqslant r$ so finally $\mu_t^{\Rightarrow}(x,y,r)$. Clearly, by definition, from $\mu_t^{\Rightarrow}(x,y,r)$ and $s < r$ it does follow that $\mu_t^{\Rightarrow}(x,y,s)$. □

We recollect here the basic cases of rough inclusions obtained from most frequently applied t–norms, cf., e.g., (Ref. 38). In all cases, $\mu_t^{\Rightarrow}(x,y,1)$ if and only if $x \leqslant y$ so the associated *ing* relation is $\leqslant$ and the underlying part relation is $<$. For $r < 1$, i.e., $x > y$, one has

Case 1. $t = L$; in this case $x \Rightarrow_L y = min\{1, 1-x+y\}$, hence $\mu_L^{\Rightarrow}(x,y,r)$ if and only if $1-x+y \geqslant r$.

Case 2. $t = P$ where $P(x,y) = x \cdot y$; in this case, $x \Rightarrow_P y = \frac{y}{x}$ when $x \neq 0$ and 1 when $x = 0$ hence $\mu_P^{\Rightarrow}(x,y,r)$ if and only if $y \geqslant x \cdot r$.

Case 3. $t = min(x,y)$; in this case $x \Rightarrow_m iny$ is y hence $\mu(x,y,r)$ if and only if $y \geqslant r$.

It has been proved (Ref. 4) that all rough inclusions induced from either non–archimedean or continuous t–norms in the manner as above are transitive in the sense of the formula: $\frac{\mu(u,v,r),\mu(v,w,s)}{\mu(u,w,t(r,s)}$.

5.2. Modifications and weaker variants of rough inclusions

In applications to be presented, some modified rough inclusions or weaker similarity measures will be instrumental, and we include a discussion of them here.

5.2.1. Modifications by means of metrics on attribute values

For the rough inclusion μ_L, the formula $\mu_L(v,u,r)$ means that $ind(v,u) \geqslant r$, i.e., at least $r \cdot 100$ percent of attributes agree on u and v; an extension of this rough inclusion depends on a chosen metric ρ bounded by 1 in the attribute value space V (we assume a simple case that ρ is defined for all attribute value sets).

Then, given an $\varepsilon \in [0,1]$, we let (Ref. 9),

$$\mu^{\varepsilon}(v,u,r) \Leftrightarrow |\{a \in A : \rho(a(v),a(u)) < \varepsilon\}| \geqslant r \cdot |A|. \tag{20}$$

it is manifest that μ^{ε} is a rough inclusion if ρ is a non–archimedean metric, i.e., $\rho(u,w) \leqslant max\{\rho(u,v),\rho(v,w)\}$, otherwise, the monotonicity condition RM 2 of sect. 5 need not be satisfied and this takes place with most popular metrics like Euclidean, Manhattan, or Minkowski's p-metrics.

5.2.2. Weak variants of rough inclusions

It is desirable to take into account also distribution of values of attributes on objects; to this end, we introduce quasi–rough inclusions which as a rule do not observe the monotonicity property RM2 (Ref. 40).

We introduce for given objects u,v, and $\varepsilon \in [0,1]$, factors: $dis_{\varepsilon}(u,v) = \frac{|\{a \in A: \rho(a(u),a(v)) \geqslant \varepsilon\}|}{|A|}$, and $ind_{\varepsilon}(u,v) = \frac{|\{a \in A: \rho(a(u),a(v)) < \varepsilon\}|}{|A|}$, where ρ is a metric on the attribute value set V bounded by 1.

Then, we modify the formula (29)to the form,

$$\nu(u,v,r) \text{ if and only if } dis_{\varepsilon}(u,v) \rightarrow_t \ ind_{\varepsilon}(u,v) \geqslant r. \tag{21}$$

Clearly, ν has properties: 1. $\nu(u,u,1)$; 2. $\nu(u,v,r)$ and $s < r$ imply $\nu(u,v,s)$ but monotonicity property RM 2 need not hold.

Rough inclusions defined above can be applied in granulation of knowledge (Ref. 41).

6. Granulation of knowledge

Formal theory of rough inclusions allows for a formal mechanism of granulation of knowledge; we assume an information system (U,A,f) given. Granulation of knowledge, proposed as a paradigm by L. A. Zadeh (Ref. 42), see also (Refs. 43, 44), means grouping objects into collections called *granules*, objects within a granule being similar with respect to a chosen measure; granular computing means computing with granules in place of objects. Into rough set theory the idea of granulation was brought in by T.Y. Lin (Ref. 45, 46) see also (Ref. 47).

The mechanism of granule formation based on rough inclusions has been presented by Polkowski in a number of works, see, e.g. (Refs. 6, 7, 9, 11, 41), and we recall it here. The basic tool in establishing properties of granules is the class operator of mereology, see sect. 4 along with the Lesniewski Inference Rule (IR), see sect. 4.

Given a rough inclusion μ on the set U, for each object u and each $r \in [0,1]$, the granule $g_{\mu}(u,r)$ of the radius r about u relative to μ is defined as the class of the property $\Phi(u,r,\mu) = \{v : \mu(v,u,r)\}$:

$$g_\mu(u,r) \ \ is \ Cls\Phi(u,r,\mu). \tag{22}$$

Thus, a granule $g_\mu(u,r)$, is constructed as the class of all objects v with the property that v is a part of u to degree at least r.

In case of t–norm–induced rough inclusions, by their transitivity, the important property holds (Ref. 5).

$$v \ ing \ g_{\mu_t}(u,r) \text{ if and only if } \mu_t(v,u,r), \tag{23}$$

i.e, the granule $g_{\mu_t}(u,r)$ is the set $\Phi(u,r,\mu_t)$.

6.1. Granular reflections of data sets

An idea of granulated data sets was proposed in (Refs. 5, 6). Given a decision system (U,A,f,d), a rough inclusion μ on the universe U, and a radius $r \in [0,1]$, one can find granules $g_\mu(u,r)$ taken in accordance to the formula (23) for all $u \in U$ and make them into the set $Gran(U,r,\mu)$. From this set, a covering $Cov(U,r,\mu)$ of the universe U can be selected by means of a strategy $\mathscr{G}$, i.e.,

$$Cov(U,r,\mu) = \mathscr{G}(Gran(U,r,\mu)). \tag{24}$$

Each granule g in $Cov(U,r,\mu)$ is a collection of objects; attributes in the set $A \cup \{d\}$ can be factored through the granule g by means of a chosen strategy $\mathscr{S}$, i.e., for each attribute $q \in A \cup \{d\}$, the new factored attribute $\overline{q}$ is defined by means of the formula,

$$\overline{q}(g) = \mathscr{S}(\{a(v) : v \in \Phi(u,r,\mu)\}). \tag{25}$$

We denote this value assignment with the symbol $\overline{f}$.

In effect, a new decision system $\mathscr{F}(U) = (Cov(U,r,\mu), \{\overline{a} : a \in A\}, \overline{d}, \overline{f})$ is defined which is called the *granular reflection of the original system*. The object v with $Inf(v) = \{(\overline{a} = \overline{a}(g)) : a \in A\}$ is called the *granular reflection of g*. Granular reflections of granules need not be objects found in data set; yet, the results show that they mediate very well between the training and test sets.

7. Granular Classification

In the sequel, we present results of tests with real data of classifiers induced from granular reflections of data as well as classifiers constructed by means of voting schemes based on variants of rough inclusions. We begin with classifiers based on the rough inclusion $\mu_h(u,v,r) = \mu_L(u,v,r)$, see equations (16), (18).

As is usual in classification tasks, the data set is split into training and test parts (sets) and we proceed in accordance with well–known procedures of *cross–validation of results*. In non–granulated case, we use the exhaustive classifier and for purpose of standardization, we apply the exhaustive classifier available in the RSES system (Ref. 23). The general procedure consists in splitting a given data set into the training and test sets, forming a granular reflection of the training set, for a given granule radius r, inducing classification rules from this new data set by the exhaustive classifier and applying the induced rules in classifying data in the test set. In case $r = 0$, the exhaustive classifier is induced from the non–granulated training set, which gives the comparison and evaluation of the effectiveness of granulation for radii in the interval $[0,1]$.

We include some examples showing the high efficiency of this approach (Refs. 40, 48, 49).

7.1. The case of $\mu_h = \mu_L$

In tests with μ_h, the strategy $\mathscr{G}$ was a random choice of an irreducible covering from the set of all granules of a given radius and the strategy $\mathscr{S}$ was chosen as majority voting with random tie resolution. Other possible strategies are, e.g.: an ordered choice for $\mathscr{G}$, a choice of granule centers for $\mathscr{S}$, etc.

We show results of tests with Australian Credit data set (Ref. 50), well studied in rough set literature: we include for comparison, some best results obtained by means of some other rough set–based methods, in Table 1. Classification quality is expressed by means of two factors: *(total) accuracy* which is the ratio of the number of correctly classified objects to the number of recognized test objects) and *(total) coverage*, $\frac{rec}{test}$, where *rec* is the number of recognized test cases and *test* is the number of test cases.

Table 1. Best results for Australian credit by some rough set based algorithms

source	*accuracy*	*coverage*	
(Ref. 14)	-	$error = 0.130$	-
(Ref. 36)	0.929	0.623	
(Ref. 36)	0.886	0.905	
(Ref. 36)	0.875	1.0	
(Ref. 51)	0.863	–	

Tests on this data with granular approach indicated above were carried by splitting the Australian credit data set into the training and test sets in the ratio of 1:1; the training sample was granulated and a granular reflection was formed from which by means of RSES exhaustive algorithm a classifier was produced which was applied to the test part of data to find quality of classification.

Granules were calculated in a twofold way: first as indicated above and second, by a modified procedure of *concept dependent granulation* (Refs. 40, 49): in the latter procedure, the granule $g_h^c(u,r) = g_h(u,r) \cap [u]_d$ was computed relative to the *concept*, i.e., decision class, to which u belonged. The results of tests are given in Table 2 in which the best results obtained with various granulation radii are shown.

Table 2. Best results for Australian credit by granular approach

source	*accuracy*	*coverage*
(Refs. 40, 48, 49)	0.867	1.0
(Refs. 40, 48, 49)	0.875	1.0

Results in Table 2 do witness that granular approach gives results fully comparable with other results for satisfactorily large radii of granulation.

In order to test the impact which the choice of granular covering has on classification, we have carried out 10 experiments with random coverings on the Heart data set (Cleveland) (Ref. 50). Results are given in Table 3.

Table 3. Effect of a choice of a granular covering on classification

radius	*total accuracy*	*total coverage*
0.0	0.0	0.0
0.0769231	0.0	0.0
0.153846	0.0	0.0
0.230769	$0.0 - 0.789$	$0.0 - 1.0$
0.307692	0.01.0	$0.0 - 1.0$
0.384615	$0.737 - 0.799$	$0.993 - 1.00$
0.461538	$0.778 - 0.822$	$0.996 - 1.0$
0.538462	$0.881 - 0.911$	1.0
0.615385	$0.874 - 0.907$	1.0
0.692308	$0.963 - 0.974$	1.0
0.769231	1.0	1.0
0.846154	1.0	1.0
0.923077	1.0	1.0
1.0	1.0	1.0

Total accuracy was found to be 0.807, and total coverage 1.0 with exhaustive algorithm on the full data. These values are achieved here with radius of at least 0.538462, and beginning with the radius of 0.384615, the error in total accuracy is at most 0.07, and the error in total coverage is at most 0.007.

This does witness a very high stability of the granular approach showing the essential independence of results of a choice of a granular covering for inducing a granular reflection of data.

7.2. The case of parameterized variants of μ_h

As discussed in (Ref. 40), 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 $\mu_h^\varepsilon(v,u,s)$ if and only if $|\{a \in A : \rho(a(v),a(u)) < \varepsilon\}| \geqslant s \cdot |A|$. The parameter s is called in this case the *catch radius*.

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

Usage of granules induced by μ_h^ε 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, $g_h^{\varepsilon}(u,1) = \{R \in Rul : \rho(a(u),a(R)) \leqslant \varepsilon$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 decision class c, the following factor is computed,

$$\text{param(c)} = \frac{\text{sum of supports of rules pointing to c}}{\text{cardinality of c in the training set}}, \tag{26}$$

cf., (Ref. 19) for a discussion of various strategies of voting for decision values.

The class c_u assigned to u is decided by

$$param(c_u) = max_c param(c), \tag{27}$$

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,

$$f_a(x) = \begin{cases} min_{training\ set} a \text{ in case } x \leqslant min_{training\ set} a \\ x \text{ in case } x \in [min_{training\ set} a, max_{training\ set} a] \\ max_{training\ set} a \text{ in case } x \geqslant max_{training\ set} a. \end{cases} \tag{28}$$

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 applied to this data set gives accuracy of 0.79 and coverage of 1.0.

7.2.1. *The case of 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.

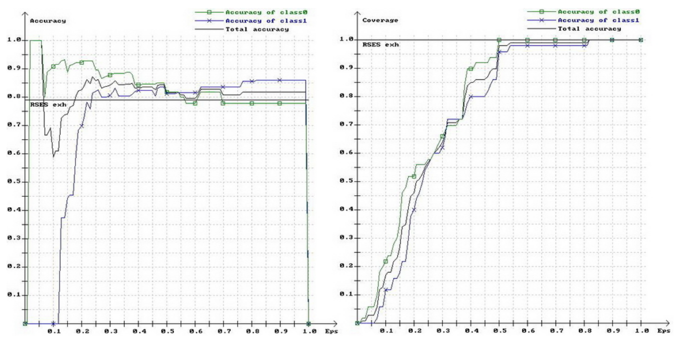

Fig. 1. Results for voting by granules of training objects. Best result for ε = 0.62: accuracy = 0.828283, coverage = 0.99

7.2.2. *The case of granules of training objects according to $\mu_h^{\varepsilon}(v,u,s)$ voting for decision*

We return to the rough inclusion $\mu_h^{*}(v,u,r)$ with general catch radius s. The procedure applied 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, catch radius.

In Table 4 results are included where against values of the catch radius s the best value for ε's marked by the optimal value *optimal eps* is given for accuracy and coverage.

The results show the effectiveness of the method: the optimal value of accuracy of 0.86 at the catch radius of 0.785714 and optimal ε of 0.39 exceeds by 0.07 the accuracy by the plain rough set exhaustive classifier.

Table 4. Results of voting by granules of training objects; 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

7.2.3. *The case of weak variants from residual implications*

As shown in (Ref. 9), residual implications of continuous t–norms can supply rough inclusions according to a general formula,

$$\mu_\phi(v,u,r) \text{ iff } \phi(u) \Rightarrow_t \phi(v) \geqslant r, \qquad (29)$$

where ϕ maps the set U of objects into $[0,1]$ and $\phi(u) \leqslant \phi(v)$ if and only if u *ing* v (*ing* is an ingredient relation of the underlying mereology, see e.g., (Ref. 9)); $\Rightarrow_t$ is the residual implication induced by the t–norm.

Candidates for ϕ have been proposed in (Ref. 9), and a weak interesting variant of this class of rough inclusions is indicated. This variant uses sets $dis_\varepsilon(u,v) = \frac{|\{a\in A:\rho(a(u),a(v))\geqslant\varepsilon\}|}{|A|}$, 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,

$$\mu_t(u,v,r) \text{ iff } dis_\varepsilon(u,v) \Rightarrow_t \; ind_\varepsilon(u,v) \geqslant r. \qquad (30)$$

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 $x \leqslant y$),

$$\mu_t(u,v,r) \text{ iff } \begin{cases} 1 - dis_\varepsilon(u,v) + ind_\varepsilon(u,v) \geqslant r \text{ for L} \\ \frac{ind_\varepsilon(u,v)}{dis_\varepsilon(u,v)} \geqslant r \text{ for P} \\ ind_\varepsilon(u,v) \geqslant r \text{ for min} \end{cases} \qquad (31)$$

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 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).

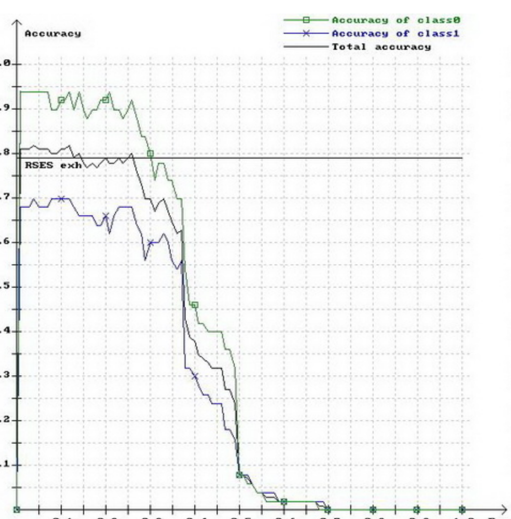

Fig. 2. Results of voting by training objects. Best result for $\varepsilon = 0.04$, accuracy = 0.82, coverage = 1

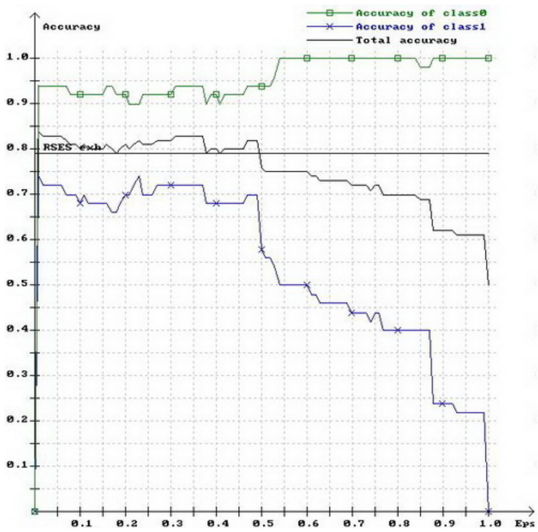

Fig. 3. Results of voting by rules. Best result for $\varepsilon = 0.01$, accuracy = 0.84, coverage = 1

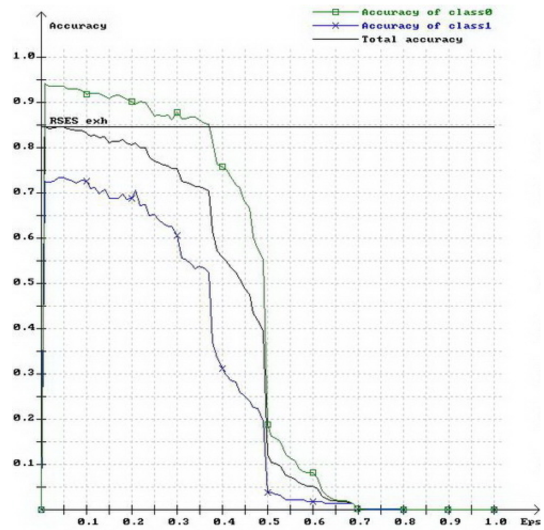

Fig. 4. Results for algorithm 5_v1, Best result for $\varepsilon = 0.04$: accuracy = 0.847826, coverage = 1

7.2.4. *The case of voting by granules*

Tests have been carried out with Australian Credit data set (Ref. 50) and the method of result validation was CV-5 (the 5–fold cross validation) (Ref. 52). Results of classification have been judged by accuracy and coverage factors. The accuracy computed with the standard RSES exhaustive classifier for these data is 0.845, and coverage is 1.0.

We have four cases for testing with: 1. granules of objects in the training set, 2. granules of rules from the training set, 3. granules of granular objects, for each of the three rough inclusions t=min, P, L.

In Case 1, training objects are made into granules for a given ε. Objects in each granule g about a test object u, vote for decision value at u as follows: for each decision class c, the value,

$$p(c) = \frac{\sum_{\text{training object v in g falling in c}} w(v,t)}{\text{size of c in training set}} \tag{32}$$

is computed where the weight $w(v,t)$ is computed for a given t–norm t as,

$$w(v,t) = dis_{\varepsilon}(u,v) \rightarrow_t ind_{\varepsilon}(u,v). \tag{33}$$

The class c* assigned to u is the one with the largest value of p. Results for the three chosen t–norms are given in Fig.4 (t=min), Fig.5 (t=P), Fig.6 (t=L).

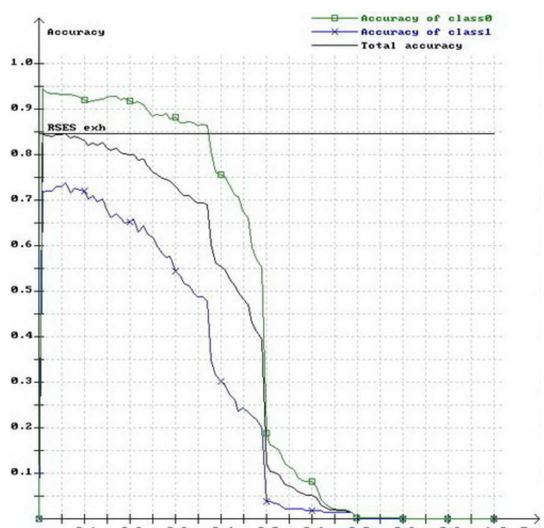

Fig. 5. Results for algorithm 6_v1, Best result for $\varepsilon = 0.06$: accuracy = 0.847826, coverage = 1

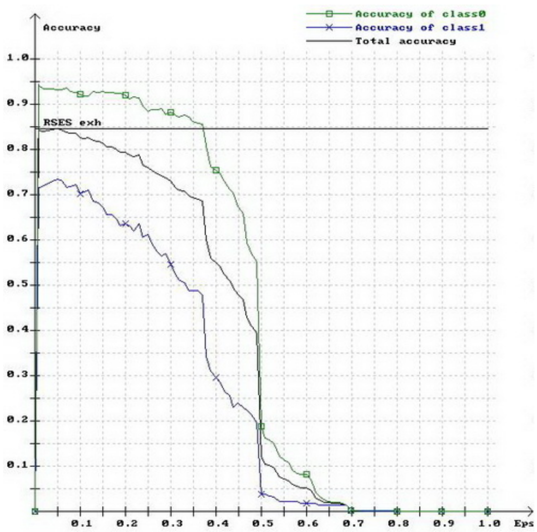

Fig. 6. Results for algorithm 7_v1, Best result for $\varepsilon = 0.05$: accuracy = 0.846377, coverage = 1

In Case 2, weighted voting of rules in a given granule g for decision at test object u goes according to the formula $d(u) = argmaxp(c)$, where

$$p(c) = \frac{\sum_{\text{rule R in g pointing to c}} w(R,t) \cdot support(R)}{\text{size of c in training set}}, \tag{34}$$

where weight is computed as,

$$w(R,t) = dis_{\varepsilon}(u,r) \rightarrow_t ind_{\varepsilon}(u,r). \tag{35}$$

Results are shown in Fig. 7 (t=min), Fig. 8 (t=P), Fig.9 (t=L).

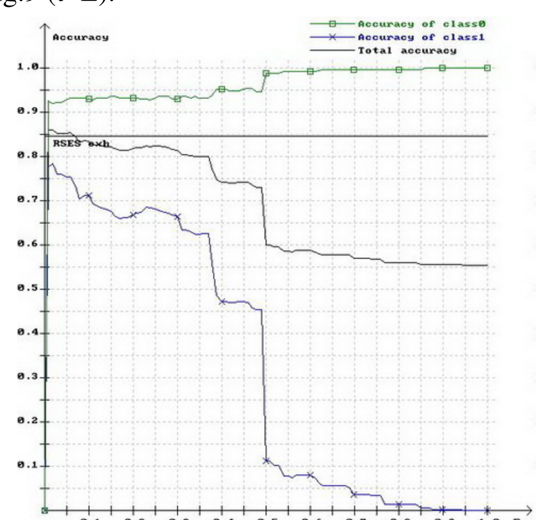

Fig. 7. Results for algorithm 5_v2, Best result for $\varepsilon = 0.02$: accuracy = 0.86087, coverage = 1

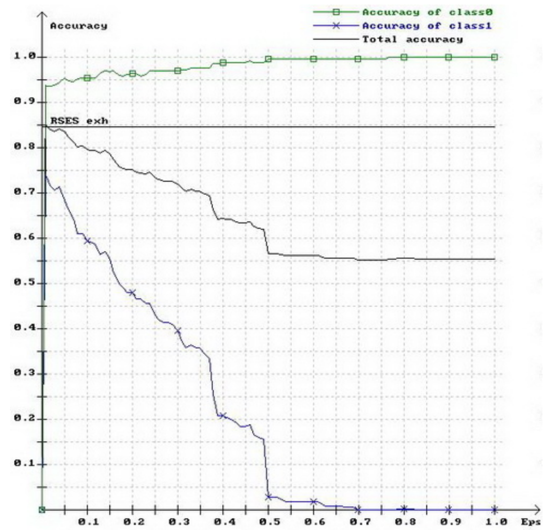

Fig. 8. Results for algorithm 6_v2, Best result for $\varepsilon = 0.01$: accuracy = 0.850725, coverage = 1

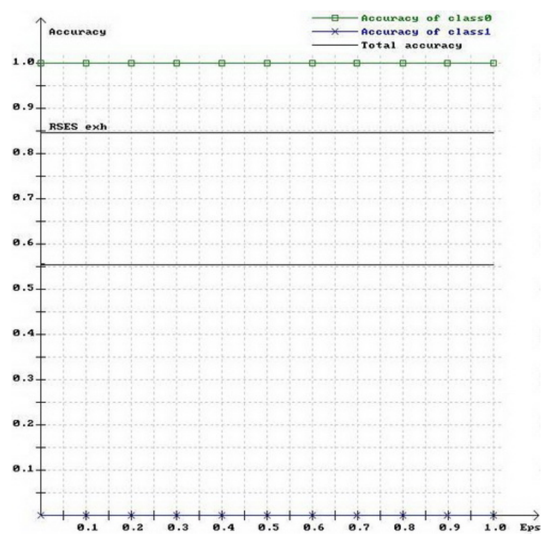

Fig. 9. Results for algorithm 7_v2, Best result for $\varepsilon = 0$, accuracy = 0.555073, coverage = 1

In Case 3, granule reflections induced, see sect. 6, from granules vote for decision. The difference is in the fact that now we have two–parameter case with ε, r hence results are given in Table 5 (t=min), Table 6 (t=P), Table 7 (t=L) in which for each row corresponding to the radius of granulation the best ε is given along with accuracy and coverage in that case.

Table 5. CV-5; Australian credit; Algorithm 5_v3. r_gran=granulation radius, optimal_eps= optimal epsilon, acc= accuracy, cov=coverage, m_trn=mean training set

r_gran	*optimal eps*	*acc*	*cov*	*m_trn*
nil	*nil*	0.845	1.0	552
0.500000	0.03	0.834783	1.0	53.8
0.571429	0.02	0.791304	1.0	134.4
0.642857	0.01	0.798551	1.0	295.8
0.714286	0.02	0.83913	1.0	454.8
0.785714	0.05	0.855072	1.0	533.8
0.857143	0.05	0.847826	1.0	546.2
0.928571	0.04	0.847826	1.0	548
1.000000	0.04	0.847826	1.0	552

Table 6. CV-5; Australian credit; Algorithm 6_v3. r_gran=granulation radius, optimal_eps= optimal epsilon, acc= accuracy, cov= coverage, m_trn=mean training set

r_gran	*optimal eps*	*acc*	*cov*	*m_trn*
nil	*nil*	0.845	1.0	552
0	0.01	0.555073	1.0	1
0.500000	0.01	0.808696	1.0	54.8
0.571429	0.01	0.746377	1.0	131.8
0.642857	0.01	0.763768	1.0	295.2
0.714286	0.01	0.818841	1.0	454.4
0.785714	0.01	0.852174	1.0	533.2
0.857143	0.01	0.847826	1.0	546.2
0.928571	0.01	0.846377	1.0	548
1.000000	0.06	0.847826	1.0	552

Table 7. CV-5; Australian credit; Algorithm 7_v3. r_gran=granulation radius, optimal_eps= optimal epsilon, acc=Total accuracy, cov=Total coverage, m_trn=mean training set

r_gran	*optimal eps*	*acc*	*cov*	*m_trn*
nil	*nil*	0.845	1.0	552
0.500000	0.01	0.707247	1.0	53.2
0.571429	0.01	0.595652	1.0	132
0.642857	0.01	0.563768	1.0	292.2
0.714286	0.02	0.786956	1.0	457.6
0.785714	0.01	0.85942	1.0	533
0.857143	0.05	0.847826	1.0	546.2
0.928571	0.05	0.849275	1.0	548
1.000000	0.05	0.846377	1.0	552

Conclusions

We have surveyed basic means for inducing rough inclusions in data sets. Applications to classification of data have been tested with real data with very good results: optimal results obtained with these classifiers are on par with best results obtained by other rough set methods, cf., Tab. 1. Optimal results obtained with granules of training objects relative to the all three t–norms and with granules of decision rules induced from the original training set for t=*min* and t=P are fully comparable with best results by rough set techniques. Comparison with classification by exhaustive classifier in non–granulated case shows that granulation heuristics enhances classification quality. In the last three tables, we show also the size of the granular reflection of the training set: here, we observe the reduction of its size in comparison to the non–granular case: for optimal classification results, the reduction is about 4-5 percent.

We mention some problems to be approached in future research in a deeper analysis of results and methods.

OPTIMAL GRANULATION RADIUS PROBLEM (OGRP)

Input: A data set

Problem: Determine optimal value r_{opt} of the granulation radius r at which the factor $accuracy \cdot coverage$ reaches the maximum value

OPTIMAL EPSILON PROBLEM (OEP)

Input: A data set, a granulation radius r

Problem: Determine optimal value ε_{opt} of the parameter ε at which the factor $accuracy \cdot\ coverage$ reaches the maximum value

References

1. Z. Pawlak, "Rough sets", *Int. J. Computer and Information Sci.*, **11**, 341–356 (1982).
2. Z. Pawlak, *Rough Sets: Theoretical Aspects of Reasoning about Data*, Kluwer, Dordrecht (1991).
3. L. Polkowski, "A rough set paradigm for unifying rough set theory and fuzzy set theory", *Lecture Notes in Artificial Intelligence*, vol. **2639**, Springer, Berlin, 70–78 (2003); cf., *Fundamenta Informaticae*, **54**, 67–88 (2003).

4. L. Polkowski, "Toward rough set foundations. Mereological approach", *Lecture Notes in Artificial Intelligence*, vol. **3066**, Springer, Berlin, 8–25 (2004).
5. L. Polkowski, "Rough–fuzzy–neurocomputing based on rough mereological calculus of granules", *International Journal of Hybrid Intelligent Systems*, **2**, 91–108 (2005).
6. L. Polkowski, "Formal granular calculi based on rough inclusions (a feature talk", *Proceedings 2005 IEEE Intern. Conf. Granular Computing GrC 2005*, IEEE Press, Piscataway NJ, 57–62 (2005).
7. 7. L. Polkowski, "Formal granular calculi based on rough inclusions (a feature talk)", *Proceedings 2006 IEEE Intern. Conf. Granular Computing GrC 2006*, IEEE Press, Piscataway NJ, 9–18 (2006).
8. L. Polkowski, "Rough mereological reasoning in rough set theory: Recent results and problems", *Lecture Notes in Artificial Intelligence*, vol. **4062**, Springer, Berlin, 79–92 (2006).
9. L.Polkowski, "Granulation of knowledge in decision systems:The approach based on rough inclusions. The method and its applications", *Lecture Notes in Artificial Intelligence*, vol. **4585**, Springer Verlag, Berlin, 69-79 (2007).
10. L. Polkowski, "Rough mereology in analysis of vagueness", *Lecture Notes in Artificial Intelligence*, vol. **5009**, Springer Verlag, Berlin, 197–204 (2008).
11. L. Polkowski, "On the idea of using granular rough mereological structures in classification of data", *Lecture Notes in Artificial Intelligence*, vol. **5009**, Springer Verlag, Berlin, 213–220 (2008).
12. F. Baader, D. Calvanese, D. McGuinness, D.Nardi, P. Patel-Schneider (Eds.), *"The Description Logic Handbook: Theory, Implementation and Applications"*, Cambridge University Press, Cambridge (2004).
13. A. Skowron, C. Rauszer, "The discernibility matrices and functions in decision systems", R. Słowiński (Ed.), *'Intelligent Decision Support. Handbook of Applications and Advances of the Rough Sets Theory',* Kluwer, Dordrecht, 311–362 (1992).
14. J. G. Bazan, Nguyen Hung Son, P. Synak, J. Wróblewski, Nguyen Sinh Hoa, "Rough set algorithms in classification problems", L. Polkowski, S. Tsumoto, T.Y. Lin (Eds.), *'Rough Set Methods and Applications. New Developments in Knowledge Discovery in Information Systems',*, Physica Verlag, Heidelberg, 49–88 (2000).
15. Z. Pawlak, A. Skowron, "A rough set approach for decision rules generation", *Proceedings of IJCAI'93 Workshop W12. The Management of Uncertainty in AI* (1993); also *ICS Research Report*, **23/93**, Warsaw University of Technology, Institute of Computer Science (1993).
16. A. Skowron, "Boolean reasoning for decision rules generation", *Lecture Notes in Artificial Intelligence*, Vol. **689**, Springer Verlag, Berlin, 295–305 (1993).
17. T. Mitchell, *"Machine learning"*, McGraw–Hill, Englewood Cliffs NJ (1997).
18. R. S. Michalski et al., "The multi–purpose incremental learning system AQ15 and its testing to three medical domains", *Proceedings of AAAI-86*, Morgan Kaufmann, San Mateo CA, 1041–1045 (1986).
19. J. G. Bazan, "A comparison of dynamic and non–dynamic rough set methods for extracting laws from decision table", L.Polkowski, A.Skowron (Eds.), *'Rough Sets in Knowledge Discovery 1',* Physica Verlag, Heidelberg, 321–365 (1998).
20. J. W. Grzymala–Busse, Ming Hu, "A comparison of several approaches to missing attribute values in Data Mining", *Lecture Notes in Artificial Intelligence* , vol. **2005**, Springer Verlag, Berlin, 378–385 (2000).
21. J. Stefanowski, "On rough set based approaches to induction of decision rules", L. Polkowski, A. Skowron(Eds.), *Rough Sets in Knowledge Discovery 1*, Physica Verlag, Heidelberg, 500–529 (1998).
22. J. W. Grzymala–Busse, "Data with missing attribute values: Generalization of indiscernibility relation and rule induction", *Transactions on Rough Sets*, I, subseries of *Lecture Notes in Computer Science*, vol. **3100**, Springer Verlag, Berlin, 78–95 (2004).
23. RSES: Available at http://logic.mimuw.edu.pl/rses
24. J. Stefanowski, "On combined classifiers, rule induction and rough sets", *Transactions on Rough Sets*, VI, a subseries of *Lecture Notes in Computer Science*, vol. **4374**, Springer Verlag, Berlin, pp 329–350 (2007).
25. Z. Pawlak, A. Skowron, "Rough membership functions", R.R.Yager, M.Fedrizzi, J.Kasprzyk (Eds.), *Advances in the Dempster–Shafer Theory of Evidence*, John Wiley and Sons, New York, 251–271 (1994).
26. H. Poincaré, *"Science and Hypothesis"*, Walter Scott Publ., London (1905).
27. E. C. Zeeman, "The topology of the brain and the visual perception", K.M. Fort (Ed.), *Topology of 3–manifolds and Selected Topics*, Prentice Hall, Englewood Cliffs, NJ, 240–256 (1965).
28. J. Nieminen, "Rough tolerance equality and tolerance black boxes", *Fundamenta Informaticae*, **11**, 289–296 (1988).
29. L. Polkowski, A.Skowron, and J.Żytkow, "Tolerance based rough sets", T.Y. Lin, M.A. Wildberger (Eds.), *Soft Computing:Rough Sets, Fuzzy Logic, Neural Networks, Uncertainty Management*, Simulation Councils, Inc., San Diego, 55–58 (1995).
30. R. Słowiński and D. Vanderpooten, "A generalized definition of rough approximations based on similarity", *IEEE Transactions on Data and Knowledge Engineering*, **12**(2), 331-336 (2000).
31. S. Leśniewski, "On the foundations of set theory", *Topoi* **2**, 7–52 (1982). See also: S. Leśniewski, *"Pod-*

stawy Ogólnej Teoryi Mnogosci" ("On the Foundations of General Set Theory", in Polish), The Polish Scientific Circle, Moscow (1916) and S. J. Surma, J. Srzednicki, D. I. Barnett, and F. V. Rickey (Eds.), *S. Leśniewski. Collected Works*, **1** , Kluwer, Dordrecht, 129-173 (1992).
32. L.Polkowski and A. Skowron, "Rough mereology: a new paradigm for approximate reasoning", *International Journal of Approximate Reasoning*, **15**(4), 333–365 (1997).
33. L. Polkowski and A. Skowron, "Grammar systems for distributed synthesis of approximate solutions extracted from experience", Gh. Paun and A.Salomaa (Eds.), *Grammatical Models of Multi–Agent Systems*, Gordon and Breach, Amsterdam, 316–333 (1999).
34. L.Polkowski and A.Skowron, "Towards an adaptive calculus of granules", L.A.Zadeh and J.Kacprzyk (Eds.), *Computing with Words in Information/Intelligent Systems*, **1**, Physica Verlag, Heidelberg, 201–228 (1999).
35. L. Polkowski and A. Skowron, "Rough mereological calculi of granules: A rough set approach to computation", *Computational Intelligence. An International Journal*, **17**(3), 472–492 (2001).
36. Nguyen Sinh Hoa, "Regularity analysis and its applications in Data Mining", L. Polkowski, S. Tsumoto and T.Y. Lin (Eds.), *Rough Set Methods and Applications. New Developments in Knowledge Discovery in Information Systems*, Physica Verlag, Heidelberg, 289–378 (2000).
37. L. Polkowski, *"Rough Sets. Mathematical Foundations"*, Physica Verlag, Heidelberg (2002).
38. P. H*á*jek, *"Metamathematics of Fuzzy Logic"*, Kluwer, Dordrecht (1998).
39. C.-H. Ling, "Representation of associative functions", *Publ. Math. Debrecen*, **12**, 189–212 (1965).
40. L. Polkowski and P. Artiemjew, "On granular rough computing: Factoring classifiers through granular structures", *Lecture Notes in Artificial Intelligence*, vol. **4585**, Springer Verlag, Berlin, 280–290 (2007).
41. L. Polkowski, 'An Approach to Granulation of Knowledge and Granular Computing Based on Rough Mereology: A Survey", W. Pedrycz, A. Skowron and V. Kreinovich (Eds.), *Handbook of Granular Computing*, John Wiley and Sons, Chichester UK (2008).
42. L. A. Zadeh, "Fuzzy sets and information granularity", M. Gupta, R. Ragade and R. Yager(Eds.), *Advances in Fuzzy Set Theory and Applications*, North–Holland, Amsterdam, 3–18 (1979).
43. L.A. Zadeh, "Toward a unified theory of uncertainty", *Proceedings IPMU 2004 (International Conference on Information Processing and Management of Uncertainty in Knowledge Based Systems)*, Editrice Univ. La Sapienza, Rome, 3-4 (2004).
44. L. A. Zadeh, "Graduation and Granulation are keys to computation with information described in Natural Language", *Proceedings of 2005 IEEE Conference on Granular Computing,GrC05*, IEEE Press, 30 (2005).
45. T.Y. Lin, "From rough sets and neighborhood systems to information granulation and computing with words", *Proceedings of the European Congress on Intelligent Techniques and Soft Computing*, 1602–06 (1997).
46. T.Y. Lin, "Neighborhood systems and relational database. Abstract", *Proceedings of CSC'88*, 725 (1988).
47. T.Y. Lin, "Granular computing: Examples, intuitions and modeling", *Proceedings of 2005 IEEE Conference on Granular Computing,GrC05*, IEEE Press, 40–44 (2005).
48. P. Artiemjew, "On classification of Data by means of rough mereological granules of objects and rules", *Lecture Notes in Artificial Intelligence*, vol. **5009**, Springer Verlag, Berlin, 221–228 (2008).
49. P. Artiemjew, "On Strategies of Knowledge Granulation with Applications to Decision Systems", PhD Dissertation, Polish–Japanese Institute of Information Technology, Warszawa, pp. 355 (2009).
50. Univ. California at Irvine Repository: Available at http://www.ics.uci.edu./ mlearn/databases/
51. J. Wróblewski, "Adaptive aspects of combining approximation spaces", S.K. Pal, L. Polkowski and A. Skowron (Eds.), *Rough Neural Computing. Techniques for Computing with Words*, Springer Verlag, Berlin, 139–156 (2004).
52. T. Hastie, R. Tibshirani and J. Friedman, *"The Elements of Statistical Learning"*, Springer Verlag, New York (2003).