Rough Sets of Zdzisław Pawlak Give New Life to Old Concepts. A Personal View ....

Lech T. Polkowski(✉)

Polish-Japanese Academy of Information Technology,
Koszykowa str. 86, 02-008 Warsaw, Poland
polkow@pjwstk.edu.pl

Abstract. Zdzisław Pawlak influenced our thinking about uncertainty by borrowing the idea of approximation from geometry and topology and carrying those ideas into the realm of knowledge engineering. In this way, simple and already much worn out mathematical notions, gained a new life given to them by new notions of decision rules and algorithms, complexity problems, and problems of optimization of relations and rules. In his work, the author would like to present his personal remembrances of how his work was influenced by Zdzisław Pawlak interlaced with discussions of highlights of research done in enliving classical concepts in new frameworks, and next, he will go to more recent results that stem from those foundations, mostly on applications of rough mereology in behavioral robotics and classifier synthesis via granular computing.

Keywords: Rough sets · Rough mereology · Granular computing · Betweenness · Mobile robot navigation · Kernel and residuum in data

1 Meeting Professor Pawlak First Time: First Problems

It was in the year 1992 and the person who contacted us was Professor Helena Rasiowa, the eminent world–renowned logician. Zdzisław asked me to create a topological theory of rough set spaces: He was eager to introduce into rough sets the classical structures; some logic and algebra already were therein. The finite case was well recognized so I followed an advice by Stan Ulam: *'if you want to discuss a finite case, go first to the infinite one'*, I considered information systems with countably infinitely many attributes. Let me sum up the essential results which were warmly welcomed by Zdzisław.

1.1 Rough Set Topology: A Context and Basic Notions

Assume given a set[1] (a *universe*) U of *objects* along with a sequence $A = \{a_n : n = 1, 2, \ldots\}$ of *attributes*;[2] without loss of generality, we may assume that

L.T. Polkowski—An invited Fellow IRSS talk.

[1] Results on topology of rough sets can be best found in author's [4].

[2] The pair IS = (U,A) will be called an *information system*; each $a_n \in A$ maps U into a set V of *possible values*.

V. Flores et al. (Eds.): IJCRS 2016, LNAI 9920, pp. 43–53, 2016.
DOI: 10.1007/978-3-319-47160-0_4

$Ind_n \subseteq Ind_{n+1}$ for each n, where $Ind_n = \{(u,v) : u,v \in U, a_n(u) = a_n(v)\}$. Letting $Ind = \bigcap_n Ind_n$, we may assume that the family $\{Ind_n : n = 1,2,\ldots\}$ *separates objects*, i.e., for each pair $u \neq v$, there is a class $P \in U/Ind_n$ for some n such that $u \in P, v \notin P$, otherwise we would pass to the quotient universe U/Ind. We endow U with some topologies.

1.2 Topologies Π_n, the Topology Π_0 and Exact and Rough Sets

For each n, the topology Π_n is defined as the partition topology obtained by taking as open sets unions of families of classes of the relation Ind_n. The topology Π_0 is the union of topologies Π_n for $n = 1,2,\ldots$. We apply the topology Π_0 to the task of discerning among subsets of the universe U[3]:

$$\text{A set } Z \subseteq U \text{ is} \Pi_0\text{-exact if } Cl_{\Pi_0} Z = Int_{\Pi_0} Z \text{ else } Z \text{ is } \Pi_0\text{-rough.} \tag{1}$$

1.3 The Space of Π_0-rough Sets is Metrizable

Each Π_0-rough set can be represented as a pair (Q,T) where $Q = Cl_{\Pi_0} X, T = U \setminus Int_{\Pi_0} X$ for some $X \subseteq U$. The pair (Q,T) has to satisfy the conditions: 1. $U = Q \cup T$. 2. $Q \cap T \neq \emptyset$. 3. If $\{x\}$ is a Π_0-open singleton then $x \notin Q \cap T$. We define a metric d_n as[4]

$$d_n(u,v) = 1 \text{ in case } [u]_n \neq [v]_n \text{ else } d_n(u,v) = 0. \tag{2}$$

and the metric d:

$$d(u,v) = \sum_n 10^{-n} \cdot d_n(u,v). \tag{3}$$

Theorem 1. *Metric topology of d is Π_0.*

We employ the notion of the *Hausdorff metric* and apply it to pairs (Q,T) satisfying 1–3 above, i.e., representing Π_0-rough sets. For pairs $(Q_1,T_1),(Q_2,T_2)$, we let

$$D((Q_1,T_1),(Q_2,T_2)) = max\{d_H(Q_1,Q_2), d_H(T_1,T_2)\} \tag{4}$$

and

$$D^*((Q_1,T_1),(Q_2,T_2)) = max\{d_H(Q_1,Q_2), d_H(T_1,T_2), d_H(Q_1 \cap Q_2, T_1 \cap T_2)\}, \tag{5}$$

where $d_H(A,B) = max\{max_{x\in A} dist(x,B), max_{y \in B} dist(y,A)\}$ is the Hausdorff metric on closed sets[5]. The main result is

Theorem 2. *If each descending sequence $\{[u_n]_n : n = 1,2,\ldots\}$ of classes of relations Ind_n has a non–empty intersection, then each D^*–fundamental sequence of Π_0–rough sets converges in the metric D to a Π_0–rough set. If, in addition, each relation Ind_n has a finite number of classes, then the space of Π_0–rough sets is compact in the metric D.*

[3] Cl_τ is the closure operator and Int_τ is the interior operator with respect to a topology τ.

[4] $[u]_n$ is the Ind_n-class of u.

[5] $dist(x,A) = min_{y\in A} d(x,y)$.

1.4 The Space of Almost Π_0-rough Sets is Metric Complete

In notation of preceding sections, it may happen that a set X is Π_n-rough for each n but it is Π_0-exact. We call such sets *almost rough sets*. We denote those sets as Π_ω-rough. Each set X of them, is represented in the form of a sequence of pairs $(Q_n, T_n) : n = 1, 2, \ldots$ such that for each n, 1. $Q_n = Cl_{\Pi_n} X, T_n = U \setminus Int_{\Pi_n} X$. 2. $Q_n \cap T_n \neq \emptyset$. 3. $Q_n \cup T_n = U$. 4. $Q_n \cap T_n$ contains no singleton $\{x\}$ with $\{x\}$ Π_n-open. To introduce a metric into the space of Π_ω-rough sets, we apply again the Hausdorff metric but in a modified way: for each n, we let $d_{H,n}$ to be the Hausdorff metric on Π_n-closed sets, and for representations (Q_n, T_n) and $(Q_n^*, T_n^*)_n$ of Π_ω-rough sets X, Y, respectively, we define the metric D' as:

$$D'(X,Y) = \sum_n 10^{-n} \cdot max\{d_{H,n}(Q_n, Q_n^*), d_{H,n}(T_n, T_n^*)\}. \tag{6}$$

It turns out that

Theorem 3. *The space of Π_ω-rough sets endowed with the metric D' is complete, i.e., each D'-fundamental sequence of Π_ω-rough sets converges to a Π_ω-rough set.*

Apart from theoretical value of these results, there was an applicational tint in them.

1.5 Approximate Collage Theorem

Consider an Euclidean space E^n along with an information system $(E^n, A = \{a_k : k = 1, 2, \ldots\})$, each attribute a_k inducing the partition P_k of E^n into cubes of the form $\prod_{i=1}^n [m_i + \frac{j_i}{2^k}, m_i + \frac{j_i+1}{2^k})$, where m_i runs over integers and $j_i \in [0, 2^k - 1]$ is an integer. Hence, $P_{k+1} \subseteq P_k$, each k. We consider *fractal objects*, i.e., systems of the form $[(C_1, C_2, \ldots, C_p), f, c]$, where each C_i is a compact set and f is an affine contracting mapping on E^n with a contraction coefficient $c \in (0, 1)$. The resulting fractal is the limit of the sequence $(F_n)_n$ of compacta, where 1. $F_0 = \bigcup_{i=1}^p C_i$. 2. $F_{n+1} = f(F_n)$. In this context, fractals are classical examples of Π_0-rough sets. Assume we perceive fractals through their approximations by consecutive grids P_k, so each F_n is viewed on as its *upper approximations* $a_k^+ F_n$ for each k[6]. As $diam(P_k) \to_{k\to\infty} 0$, it is evident that the symmetric difference $F \triangle F_n$ becomes arbitrarily close to the symmetric difference $a_k^+ F \triangle a_k^+ F_n$. Hence, in order to approximate F with F_n it suffices to approximate $a_k^+ F$ with $a_k^+ F_n$. The question poses itself: what is the least k which guarantees for a given ε, that if $a_k^+ F_n = a_k^+ F$ then $d_H(F, F_n) \leq \varepsilon$. We consider the metric D on fractals and their approximations. We had proposed a counterpart to Collage Theorem, by replacing fractals F_n by their grid approximations[7].

---

[6] This theorem comes from the chapter by the author in [3].

[7] The upper approximation of a set $X \subseteq U$ with respect to a partition P on U is $\bigcup\{q \in P : q \cap X \neq \emptyset\}$.

Theorem 4 *(Approximate Collage Theorem). Assume a fractal F generated by the system $(F_0 = \bigcup_{i=1}^{p} C_i, f, c)$ in the space of Π_0-rough sets with the metric D. In order to satisfy the requirement $d_H(F, F_n) \leq \varepsilon$, it is sufficient to satisfy the requirement $a^+_{k_0} F_n = a^+_{k_0} F$ with $k_0 = \lceil \frac{1}{2} - log_2 \varepsilon \rceil$ and $n \geq \lceil \frac{log_2[2^{-k_0+\frac{1}{2}} \cdot K^{-1} \cdot (1-c)]}{log_2 c} \rceil$, where $K = d_H(F_0, F_1)$.*

2 Mereology and Rough Mereology

It was a characteristic feature of Professor Pawlak that He had a great interest in theoretical questions. He remembered how He browsed through volumes in the Library at Mathematical Institute of the Polish Academy of Sciences. No doubt that the emergence of rough set theory owes much to those excursions into philosophical writings of Frege, Russell and others. At one time, Zdzisław mentioned some fascicles of the works of Stanisław Leśniewski, the creator of the first formal theory of Mereology. Zdzisław was greatly interested in various formalizations of the idea of a concept and in particular in possible relations between Mereology and Rough Sets. From our analysis of the two theories Rough Mereology emerged.

2.1 Basic Mereology

The primitive notion is here that of a *part*. The relation of *being a part of*, denoted $prt(u, v)$, is defined on a universe U by requirements: 1. $prt(u, u)$ holds for no u. 2. $prt(u, v)$ and $prt(v, w)$ imply $prt(u, w)$: $prt(u, v)$ means that u is a *proper part* of v. To account for *improper parts*, i.e., *wholes* the notion of *an ingredient, element*, *ingr* for short, was proposed which is $prt \cup$ '=', i.e., $ingr(u, v)$ if and only if $prt(u, v)$ or $u = v$. Ingredients are essential in mereological reasoning by the Leśniewski Inference Rule (LIR for short):[8]

LIR: For $u, v \in U$, if for each w such that $ingr(w, u)$, there exist t, q such that $ingr(t, w), ingr(t, q), ingr(q, v)$, then $ingr(u, v)$.

Ingredients are instrumental in forming individuals–classes of individuals: for each non-void property $\mathcal{C}$ of individuals in U, there exists a unique individual, the class of $\mathcal{C}$, $Cls\mathcal{C}$ in symbols, defined by requirements: 1. If u satisfies $\mathcal{C}$ then $ingr(u, Cls\mathcal{C})$. 2. For each u with $ingr(u, Cls\mathcal{C})$, there exist t, q such that $ingr(t, u), ingr(t, q)$ and q satisfies $\mathcal{C}$. Classes are instrumental in our definition of granules. The favorite example of Leśniewski was the chessboard as the class of white and black squares.

2.2 Rough Mereology

The basic notion of a part to a degree is rendered as the relation $\mu(u, v, r) \subseteq U^2 \times [0, 1]$, read as *'u is a part of v to a degree of at least r'* which is defined by

[8] To acquaint oneself with this theory it is best to read Lesniewski [2]. This is a rendering by E. Luschei of the original work *Foundations of Set Theory*. Polish Scientific Circle. Moscow 1916.

requirements: 1. $\mu(u, v, 1)$ if and only if $ingr(u, v)$. 2. If $\mu(u, v, 1)$ and $\mu(w, u, r)$ then $\mu(w, v, r)$. 3. If $\mu(u, v, r)$ and $s < r$ then $\mu(u, v, s)$. The relation μ was termed by us a *rough inclusion*. Relation of rough mereology to rough set theory becomes clear when we realize that the latter is about concepts and their approximations and that the containment relation is a particular case of the part relation, hence approximations upper and lower are classes of indiscernibility classes which are ingredients or, respectively, parts to a positive degree of a concept. Rough inclusions in information systems are usually defined in the attribute–value format, examples are for instance given by t–norms. It is well–known that Archimedean t-norms, the Łukasiewicz t–norm $L(x, y) = max\{0, x + y - 1\}$ and the Menger (product) t–norm $P(x, y) = x \cdot y$, allow the representation of the form $T(x, y) = g(f(x) + f(y))$, where $f : [0, 1] \to [0, 1]$ is a decreasing continuous function with $f(1) = 0$ and g is the pseudo–inverse to f. For an information system $IS = (U, A)$, the *discernibility set* $Dis(u, v)$ equals $A \setminus Ind(u, v)$[9].

Theorem 5. *For an Archimedean t–norm $T(x, y) = g(f(x) + f(y))$, the relation $\mu_T(u, v, r)$ if and only if $g(\frac{card(Dis(u,v))}{card(A)}) \geq r$ is a rough inclusion on the universe U.*

As an example, we define the *Łukasiewicz rough inclusion* μ_L as $\mu_L(u, v, r)$ if and only if $g(\frac{card(Dis(u,v))}{card(A)}) \geq r$. As in case of Łukasiewicz rough inclusion, $g(x) = 1 - x$, we have $\mu_L(u, v, r)$ if and only if $\frac{card(Ind(u,v))}{card(A)} \geq r$: a fuzzified indiscernibility. We recall that each t–norm T defines the *residual implication* $\to_T$ via the equivalence $x \to_T y \geq r$ if and only if $T(x, r) \leq y$.

Theorem 6. *Let $\to_T$ be a residual implication and $f : U \to [0, 1]$ an embedding of U into the unit interval. Then $\mu(u, v, r)$ if and only if $f(u) \to_T f(v) \geq r$ is a rough inclusion.*

We have therefore a collection of rough inclusions to be selected.

3 Rough Mereology in Behavioral Robotics

Autonomous robots are one of the best examples for the notion of an intelligent agent. Problems of their navigation in environments with obstacles are basic in behavioral robotics. We recall here an approach based on rough mereology[10].

3.1 Betweenness Relation in Navigating of Teams of Intelligent Agents

Betweenness relation is one of primitive, apart from equidistance, relations adopted by Alfred Tarski in His axiomatization of plane geometry. This relation was generalized by Johan van Bentham in the form of the relation $B(x, y, z)$,

9 Please see relevant chapters in Polkowski [5].
10 Please see Polkowski L., Osmialowski P. [8].

x, y, z points in an Euclidean space of a finite dimension (it reads: '*x is between y and z*'), with a metric d, in the form:

$$B(x, y, z) \text{ if and only if for each } q \neq x: \; d(x, y) < d(q, y) \text{ or } d(x, z) < d(q, z). \tag{7}$$

Rough mereology offers a quasi–distance function:

$$\kappa(x, y) = min\{sup_r \mu(x, y, r), sup_s \mu(y, x, s)\}. \tag{8}$$

We apply in definition of $\kappa(x, y)$ the rough inclusion $\mu(a, b, r)$, where a, b are bounded measurable sets in the plane,

$$\mu(a, b, r) \text{ if and only if } \frac{area(a \cap b)}{area(a)} \geq r. \tag{9}$$

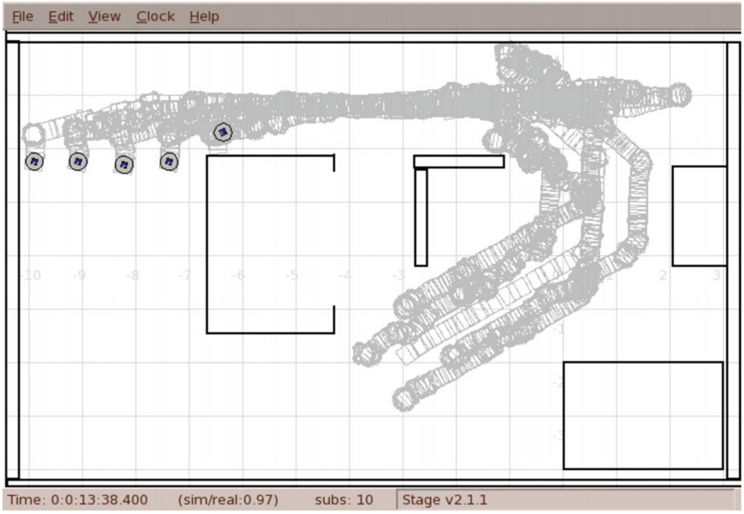

Fig. 1. Trails of robots moving in the line formation through the passage.

Consider autonomous robots in the plane as embodiments of intelligent agents. We model robots as rectangles (in fact squares) regularly placed, i.e., with edges parallel to coordinate axes. For such robots denoted a, b, c,... , the betweenness relation can be expressed as follows, see [8]:

Theorem 7. *Robot a is between robots b and c, i.e. $B(a, b, c)$ holds true, with respect to betweenness defined in (7), distance defined in (8) and the rough inclusion defined in (9) if and only if $a \subseteq ext(b, c)$, where $ext(b, c)$ is the* extent *of b and c, i.e., the minimal rectangle containing b and c.*

For a team of robots $\mathcal{T} = \{a_1, a_2, \ldots, a_m\}$, a *formation* on $\mathcal{T}$ is a relation B on $\mathcal{T}$. Figure 1 shows a team of robots in Line formation mediating a bottleneck passage after which they return to the Cross formation.

4 Granular Computing

The last of rough mereology applications Zdzisław could be acquainted with is a theory of granular computing presented first at GrC 2005 at Tsinghua University in Beijing, China. Given a rough inclusion μ on a universe U of an information system (U, A), define a *granule* $g_\mu(u, r)$ about $u \in U$ of the radius r as $g_\mu(u, r) = Cls\{v \in U : \mu(v, u, r)\}$. For practical reasons, we compute granules as sets $\{v \in U : \mu(v, u, r)\}$. The class and the set coincide for many rough inclusions, cf. [5][11].

4.1 Granular Classifers: Synthesis via Rough Inclusions

We assume that we are given a decision system $DS = (U, A, d)$ from which a classifier is to be constructed; on the universe U, a rough inclusion μ is given, and a radius $r \in [0, 1]$ is chosen. We can find granules $g_\mu(u, r)$ for all $u \in U$, and make them into the set $G(\mu, r)$. From this set, a covering $Cov(\mu, r)$ of the universe U can be selected by means of a chosen strategy $\mathcal{G}$, i.e.,[12]

$$Cov(\mu, r) = \mathcal{G}(G(\mu, r)). \tag{10}$$

We intend that $Cov(\mu, r)$ becomes a new universe of the decision system whose name will be the *granular reflection* of the original decision system. It remains to define new attributes for this decision system. Each granule g in $Cov(\mu, r)$ 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 $\mathcal{S}$, usually the majority vote, i.e., for each attribute $a \in A \cup \{d\}$, the new factored attribute $\overline{a}$ is defined by means of the formula

$$\overline{a}(g) = \mathcal{S}(\{a(v) : ingr(v, g_\mu(u, r))\}) \tag{11}$$

In effect, a new decision system $(Cov(\mu, r), \{\overline{a} : a \in A\}, \overline{d})$ is defined. The thing[13] $\overline{v_g}$ with the *information set* $Inf(\overline{v_g})$ defined as[14]

$$Inf(\overline{v_g}) = \{(\overline{a}, \overline{a}(g)) : a \in A \cup \{d\}\} \tag{12}$$

is called the *granular reflection of g*. We consider a standard data set *the Australian Credit Data Set* from Repository at UC Irvine and we collect the best results for this data set by various rough set based methods in the table of Fig. 2. In Fig. 3, we give for this data set the results of exhaustive classifier on granular structures: meanings of symbols are r = granule radius, tst = test set size, trn = train set size, rulex = rule number, aex = accuracy, cex = coverage[15].

[11] Please consult Polkowski [5] Ch. 9 and Polkowski, Artiemjew [6].

[12] An information system ***IS = (U,A)*** augmented by a new attribute $\boldsymbol{d : U \to V}$, the *decision*, is called the *decision system* ***DS = (U,A,d)***.

[13] The philosophical term ‘thing’ is reserved for beings of virtual character possibly not present in the given information/decision system.

[14] In a decision system (U, A, d), for $u \in U$, the information set of u is $Inf(u) = \{(a, a(u)) : a \in A \cup \{d\}\}$.

[15] MI is the *Michalski index*. $MI = \frac{1}{2} \cdot aex + \frac{1}{4} \cdot aex^2 + \frac{1}{2} \cdot cex - \frac{1}{4} \cdot aex \cdot cex$.

source	*method*	*accuracy*	*coverage*	*MI*
Bazan	$SNAPM(0.9)$	$error = 0.130$	$--$	$--$
Nguyen SH	*simple.templates*	0.929	0.623	0.847
Nguyen SH	*general.templates*	0.886	0.905	0.891
Nguyen SH	*tolerance.gen.templ.*	0.875	1.0	0.891
Wroblewski	*adaptive.classifier*	0.863	$-$	$--$

Fig. 2. Best results for Australian credit by some rough set based algorithms

r	*tst*	*trn*	*rulex*	*aex*	*cex*	*MI*
nil	345	345	5597	0.872	0.994	0.907
0.0	345	1	0	0.0	0.0	0.0
0.0714286	345	1	0	0.0	0.0	0.0
0.142857	345	2	0	0.0	0.0	0.0
0.214286	345	3	7	0.641	1.0	0.762
0.285714	345	4	10	0.812	1.0	0.867
0.357143	345	8	23	0.786	1.0	0.849
0.428571	345	20	96	0.791	1.0	0.850
0.5	345	51	293	0.838	1.0	0.915
0.571429	345	105	933	0.855	1.0	0.896
0.642857	345	205	3157	0.867	1.0	0.904
0.714286	345	309	5271	0.875	1.0	0.891
0.785714	345	340	5563	0.870	1.0	0.890
0.857143	345	340	5574	0.864	1.0	0.902
0.928571	345	342	5595	0.867	1.0	0.904

Fig. 3. Australian credit granulated

We can compare results: for template based methods, the best MI is 0.891, for exhaustive classifier (r = nil) MI is equal to 0.907 and for granular reflections, the best MI value is 0.915 with few other values exceeding 0.900. What seems worthy of a moment's reflection is the number of rules in the classifier. Whereas for the exhaustive classifier (r = nil) in non–granular case, the number of rules is equal to 5597, in granular case the number of rules can be surprisingly small with a good MI value, e.g., at $r = 0.5$, the number of rules is 293, i.e., 5 percent of the exhaustive classifier size, with the best MI of all of 0.915. This compression of classifier seems to be the most impressive feature of granular classifiers.

5 Betweenness Revisited in Data Sets

We can use in a given information set $IS = (U, A)$, the Łukasiewicz rough inclusion μ_L in order to obtain the mereological distance κ of (8) and the *generalized betweenness* relation GB (read: *'u is between* $v_1, v_2, ..., v_k$)[16]:

[16] A detailed account please find in Polkowski, Nowak [7].

$$GB(u, v_1, v_2, ..., v_k) \text{ if for each } v \neq u, \text{ there is } v_i \text{ such that } \kappa(u, v_i) \geq \kappa(u, v). \quad (13)$$

One proves cf. [7] that betweenness GB can be expressed as a convex combination:

Theorem 8. *$GB(u, v_1, v_2, ..., v_k)$ if and only if $Inf(u) = \bigcup_{i=1}^{k} C_i$, where $C_i \subseteq Inf(v_i)$ for $i = 1, 2, ..., k$ and $C_i \cap C_j = \emptyset$ for each pair $i \neq j$.*

In order to remove ambiguity in representing u, we introduce the notion of a *neighborhood* $N(u)$ over a set of *neighbors* $\{v_1, v_2, \ldots, v_k\}$ as the structure of the form:

$$< (v_1, C_1 \subseteq Ind(u, v_1), q(v_1)), \ldots, (v_k, C_k \subseteq Ind(u, v_k), q(v_k)) > \quad (14)$$

with neighbors $v_1, v_2, \ldots, v_k$ ordered in the descending order of the factor q, where $q_i = \frac{card(C_i)}{card(A)}$. Clearly, $\sum_{i=1}^{k} q_i = 1$ and $q_i > 0$ for each $i \leq k$.

5.1 Dual Indiscernibility Matrix, Kernel and Residuum

Dual indiscernibility matrix DIM, for short, is defined as the matrix $M(U, A) = [m_{a,v}]$ where $a \in A, v$ a value of a and $m_{a,v} = \{u \in U : a(u) = v\}$ for each pair a, v. The *residuum* of the information system (U, A), *Res* in symbols, is the set $\{u \in U$: there exists a pair(a, v) with $m_{a,v} = \{u\}\}$. The difference $U \setminus Res$ is the *kernel*, Ker in symbols. Clearly, $U = Ker \cup Res$, $Ker \cap Res = \emptyset$. The rationale behind those notions is that Ker consists of objects mutually exchangeable so averaged decisions on neighbors should transfer to test objects, while *Res* consists of objects with outliers which may serve as antennae catching test objects. It is interesting to see how those subsets do in tasks of classification into decision classes. Figure 4 shows results of applying C4.5 and k-NN to whole data set, Ker and Res for a few data sets from UC Irvine Repository. Results are very satisfying in terms of accuracy and size of data sets. Please observe that, for data considered, sets Ker and Res as a rule yield better of results for C4.5 and k-NN on the whole set[17].

5.2 A Novel Approach: Partial Approximation of Data Objects

The Pair classifier approaches a test object with inductively selected pairs of neighbors of training objects covering it partly[18].

Induction is driven by degree of covering from maximal down to the threshold number of steps. Successive pairs are indexed with *level* L. Objects in pairs up to a given level are pooled and they vote for decision value by majority voting.

---

[17] In order to split the data set into parts of which one is GB-self-contained and the other GB-vacuous, we propose the DIM matrix.

[18] A relaxed idea of convex combinations of objects lies in approximating only parts of data objects with training objects, see Artiemjew, Nowak, Polkowski [1].

database	set tested	accuracy of C4.5	accuracy of k-NN	number of samples
adult	whole set	$.857 \pm .003$	$.837 \pm .003$	39074.0
	Ker	$.853 \pm .004$	$.835 \pm .003$	22366.0
	Res	$.849 \pm .003$	$.833 \pm .003$	16708.0
PID	whole set	$.733 \pm .027$	$.723 \pm .021$	614.4
	Ker	$.704 \pm .037$	$.711 \pm .032$	212.9
	Res	$.724 \pm .035$	$.745 \pm .030$	401.5
fertility diagnosis	whole set	$.852 \pm .073$	$.866 \pm .060$	80.0
	Ker	$.846 \pm .075$	$.880 \pm .064$	71.6
	Res	$.852 \pm .068$	$.880 \pm .064$	8.4
german credit	whole set	$.713 \pm .023$	$.732 \pm .025$	800.0
	Ker	$.671 \pm .045$	$.714 \pm .038$	98.9
	Res	$.712 \pm .023$	$.726 \pm .030$	701.1
heart disease	whole set	$.750 \pm .054$	$.825 \pm .048$	216.0
	Ker	$.742 \pm .061$	$.822 \pm .051$	109.2
	Res	$.767 \pm .054$	$.827 \pm .041$	106.8

Fig. 4. Classification results

database	kNN	Bayes	Pair–best	Pair-0
Adult	.841	.864	.853L1	.823
Australian	.855	.843	.859L4,5	.859
Diabetes	.631	.652	.721L0	.710
German credit	.730	.704	.722L1	.721
Heart disease	.837	.829	.822L1	.800
Hepatitis	.890	.845	.892L0	.831
Congressional voting	.938	.927	.928L0	.928
Mushroom	1.0	.910	1.0L0	1.0
Nursery	.578	.869	.845L0	.845
Soybean large	.928	.690	.910L0	.910

Fig. 5. Pair classifier

Figure 5 shows results in comparison to k-NN and Bayes classifiers. The symbol Lx denotes the level of covering, Pair-0 is the simple pair classifier with approximations by the best pair and Pair–best denotes the best result over levels studied.

6 Conclusions

The paper presents some results along two threads: along one thread results are highlighted obtained by following Zdzisław Pawlak's 'research requests' and the other thread illustrates results obtained in classical settings by considering new contexts of knowledge engineering created by vision of Zdzisław Pawlak. Further work will focus on rational search for small decision-representative subsets of data with Big Data on mind and rough set based Approximate Ontology in biological and medical data.

Acknowledgements. This is in remembrance of Prof. Zdzisław Pawlak on the 10th anniversary of His demise. To organizers of IJCRS 2016 Prof. Richard Weber and Dr. Dominik Ślęzak DSc my thanks go for the invitation to talk henceforth the occasion to remember. To referees my thanks go for their useful remarks.

References

1. Artiemjew, P., Nowak, B., Polkowski, L.: A classifier based on the dual indiscernibility matrix. In: Dregvaite, G., Damasevicius, R. (eds.) Forthcoming in Communications in Computer and Information Science, Proceedings ICIST 2016, CCIS639, pp. 1–12. Springer (2016). doi:10.1007/978-3-319-46254-7_30
2. Lesniewski, S.: On the foundations of mathematics. Topoi **2**, 7–52 (1982)
3. Polkowski, L.: Approximate mathematical morphology. In: Pal, S.K., Skowron, A. (eds.) Rough Fuzzy Hybridization, pp. 151–162. Springer, Singapore (1999)
4. Polkowski, L.T.: Rough Sets. Mathematical Foundations. Springer, Physica, Heidelberg (2002)
5. Polkowski, L.T.: Approximate Reasoning by Parts. An Introduction to Rough Mereology. Springer, Switzerland (2011)
6. Polkowski, L., Artiemjew, P.: Granular Computing in Decision Approximation. An Application of Rough Mereology. Springer, Switzerland (2015)
7. Polkowski, L., Nowak, B.: Betweenness, Łukasiewicz rough inclusion, Euclidean representations in information systems, hyper-granules, conflict resolution. Fundamenta Informaticae **147**(2-3) (2016)
8. Polkowski, L., Osmialowski, P.: Navigation for mobile autonomous robots and their formations. An application of spatial reasoning induced from rough mereological geometry. In: Barrera, A. (ed.) Mobile Robots Navigation, pp. 329–354. Intech, Zagreb (2010)