PROBABILISTIC APPROXIMATION UNDER INCOMPLETE INFORMATION SYSTEMS

Yucai Feng, Wenhai Li, Zehua Lv and Xiaoming Ma
Department of Computer Science,
Huazhong University of Science and Technology,
Wuhan 430074, Hubei, China,
lwhaymail@21cn.com

Abstract By applying the probability estimation of the unavailable attributes derived from the available attributes to the neighborhood system, the suited degree of each neighbor to a given object is depicted. Therefore, the neighborhood space with guaranteed suited precision is obtained. We show how to shrink the rule search space via VPRS model for this space, and also, we will prove the incredibility degree of decision class is guaranteed by the two-layer thresholds.

Keywords: threshold, approximation, neighborhood system, probabilistic, rough set

1. Introduction

Classical rough sets theory [1] is too rigid to be applied to the real-life environment due to the requirement that all the characters of an object in the system are available [2]. Kryszkiewicz [4] extended the rough set approximation method in incomplete information systems through the tolerance relation with the "missing values" semantics. Slowinski [5] and Stefanowski [6] used similarity relation instead of indiscernibility relation to express the "absent values" semantics. Furthermore, the latter defined a "tolerance class" under the hypothesis that an equivalent probability was associated with each element among such values. The variable precision rough set model [7] classified the objects, probably bearing a family of misclassification(based on a β threshold) in the graduation layer [9-11].

Our research gives a probabilistic angle of view to incomplete values while generating the approximations: the distribution of "missing" or "absent" values is quantificationally taken into account when the relevant neighborhood is generated, to achieve which the guarantee of the covering quality is elucidated by a granulation threshold λ. After this, a symmetric graduation threshold β is proposed to satisfy predefined certainty requirements. The (λ, β) threshold pair includes the certain defined lines which are applied to qualify the approximation regions with the price of controllable imprecision.

Please use the following format when citing this chapter:

Feng, Y., Li, W., Lv, Z., Ma, X., 2006, in IFIP International Federation for Information Processing, Volume 228, Intelligent Information Processing III, eds. Z. Shi, Shimohara K., Feng D., (Boston: Springer), pp. 73–80.

2. Rough sets preliminaries

An information system AS is a pair (U, A, V, f) where $U \neq \emptyset$ is a non empty set of objects and A is a non empty set of attributes. For $\forall c \in A$ under $f : U \rightarrow V_c$, V_c is the domain of c and $V = \bigcup\{V_c : \forall c \in A\}$. For any $B \subseteq A$, an indiscernibility relation $IND(B)$ is defined with $IND(B) = \{(x, y) \in U \times U : \forall c_i \in B, c_i(x) = c_i(y)\}$. $U/IND(B)$(or U/B) is the family of all the equivalence classes of the indiscernibility relation $IND(B)$ with U.

For any non empty subset of objects $X \subseteq U$, the $B - lower$, $B - upper$ approximations and $B - boundary$ are defined with $X_B = \bigcup\{Y \in U/B | Y \subseteq X\}$, $X^B = \bigcup\{Y \in U/B | Y \cap X \neq \emptyset\}$, $Bn(X) = X^B \backslash X_B$ apart. The $B - lower$ is also called positive region, while the supplement of the $B - upper$ is also called negative region.

A decision information system DS is an information system while $A = C \bigcup\{d\}$ and $d \notin C$, here d is called decision and C is called condition attribute set. Any decision rule is represented as:

$$\sigma = \wedge_{c_j \in B}(c_j(a_i) = v) \rightarrow (d(a_i) = w), \tag{1}$$

where $c_j \in B \subseteq C$, and v and w are the corresponding attribute values of $a_i \in U$ respectively. Here, the left side of the implication is noted by s and t on the opposite. Let $[s]$ denote the set of the objects in DS satisfying s and $[t]$ for t accordingly. A decision rule with $s \rightarrow t$ is certain if and only if $[s] \subseteq [t]$. Otherwise, possible decision rules in the inconsistent decision information system can be induced from the upper approximation of a decision class expressed in t.

3. Probabilistic Approximation Space

AS is an incomplete information system when some values are not available, and the unavailable values are denoted by " $*$ ". The approximate relations are proposed mainly based on two hypothetical semantics for the unknown values. The lower and upper approximations are derived from the cover of relation τ instead of the indiscernibility class. The cover in a neighborhood system (U, IC) is defined as:

$$C = \bigcup\{I(x) | \forall_{x \in U, y \in I(x)}(y \tau x)\}, \tag{2}$$

in which, accordingly, $I(x)$ denotes the neighborhood of x and τ denotes the relation defined in the information system. The tolerance and similarity approach are proposed in terms of a different explanation for the unavailable values. The key concept of tolerance approach is the tolerance class. Tolerance class denoted by $I_B(x)$ of any object $x \in U$ is induced from the tolerance relation T_B which is reflexive and symmetric, but not necessarily transitive relation and obeys a "missing value" hypothesis[4]. Given an information system AS and a subset of attributes $B \subseteq A$ is defined as:

$$\forall_{x,y \in U \times U}(T_B(x, y) \Leftrightarrow \forall_{c_j \in B}(c_j(x) = c_j(y) \vee c_j(x) = * \vee c_j(y) = *)), \tag{3}$$

where for each object the binary relation T_B identifies a tolerant class $I_B(x) = \{y \in U | T_B(x, y)\}$, and consequently, the $B - lower$ and $B - upper$ of an object set are

$X_B^T = \{x \in U | I_B(x) \subseteq X\}$ and $X_T^B = \{x \in U | I_B(x) \cap X \neq \emptyset\} = \bigcup\{x \in X | I_B(x)\}$ respectively.

The non-symmetric similarity relation is similar with the tolerance relation except that the former is a partial order on the set U under an "absent value" hypothesis[5, 6]. Homoplastically, for a given information system AS and a subset of attributes $B \subseteq A$, the similarity relation S_B is defined as:

$$\forall_{x,y \in U \times U}(S_B(x,y) \Leftrightarrow \forall_{c_j \in B}(c_j(x) = * \vee c_j(x) = x_j(y))). \tag{4}$$

Consequently, similarity class $R_B(x)$ and converse similarity class $R_B^{-1}(x)$ of object $x \in U$ with respect to B are defined as $R_B(x) = \{y \in U | S_B(y,x)\}$ and $R_B^{-1}(x) = \{y \in U | S_B(x,y)\}$. $B - lower$ and $B - upper$ are $X_B^S = \{x \in U | R_B^{-1}(x) \subseteq X\}$ and $X_S^B = \bigcup\{x \in X | R_B(x)\}$ separately.

Most of the approximation methods are built based on both relations mentioned above. Furthermore, many quantitative and qualitative extensions[8] are applied to the above relations. Among the proposed methods of the incomplete information system, the unbending matching of either tolerance or similarity relation cannot control the inflation of the neighborhood, which results in the bilateral expansion of the $B - boundary$. All these induce the inefficiency of the reducing search space. From the perspective of the probability [12] of the unknown attribute value for two given objects x and y, the tolerance relation supports that any attribute value is suited iff all the available values are suited, while the valued tolerance relation regulates the possibility of the "missing value" followed by an equiprobable distribution hypothesis throughout the domain. Nevertheless, the probability of $c_j(y)$ matching $c_j(x)$ on any $c_j \in B$ with $c_j(x) = * \vee c_j(x) = *$ held is intuitively higher than $c_j(y) = 1/|V_c|$ when the available values are all suited.

DEFINITION 1 *Given $c_j \in B$ and $x, y \in U$ in an incomplete information system AS, if $c_j(x) \in V_c$ and $c_j(y) = *$, the probability of $c_j(y) \Re c_j(x)$ ($\Re$ represents the matching of y to x on c_j) denoted by $P^{\models}(c_j)$ is relative to the cardinality of all the suited attributes $|\{\forall c_j | c_j(x) \Re c_j(y) \in V_c\}|$ independently.*

Obviously, AS is in tolerance relation when $P^{\models}(c_j) = 1$ and $\Re$ denotes " $=$ " and it is in valued tolerance relation while $P^{\models}(c_j) = 1/|V_c|$. Definition 1 is given to depict the suited possibility of two objects on unknown attributes relevant to the cardinality percentage of the available suited attributes, and the unknown attributes are independent from each other.

PROPOSITION 2 *Given a neighborhood system (U, IC), the expected function of any two objects $x, y \in U$ can be expressed by the percentage of suited attributes:*

$$P(x \tau y) = \prod_{c_j(x)=* \vee c_j(y)=*}^{\forall c_j \in B} P^{\models}. \tag{5}$$

For the neighborhood system in Proposition 2, the mapping combination of x and y on c_j has three possible cases due to the unavailable values. When $c_j(y) = c_j(x) \neq *$,

the matching of y to x on c_j is certain, so the percentage of $|\{c_j(y) = c_j(x) \neq *\}|$ to $|C|$ represents the conditional probability of any uncertain matching concerning certain matching. Under the "missing" semantics, the probability of $c_j(y)\Re c_j(x)$ equals to "1" iff y are certainly suited to x on c_j, so that the total probability of all the certain suited attributes equals to "1" and is denoted by $P^{\equiv}$, otherwise, it equals to "0". When $c_j(y) = * \wedge c_j(x) \neq *$, from Definition 1, we assert that $P(c_j(y)\Re c_j(x))$ is a joint probability on $P^{\models}$ and $P^{\equiv}$, this equation comes into existence when $c_j(x) = * \wedge c_j(y) \neq *$ due to the symmetry of the tolerance relation. Furthermore, $c_j(y) = c_j(x) = *$ has a bilateral effect on the total probability of two objects. From all the above, the probability of two objects with $x\tau y$ held depends on the $P^{\models}$ exponentially, and the power is the arisen times of $*$. The similarity relation holds Proposition 2 except for the range of the percentage and the power due to its unilateralism.

The threshold λ is used to control the power of the total probability. Therefore, the neighborhood of a given object can be controlled, and accordingly, the probability of all objects in the neighborhood of the given objects is not less than $(1-\lambda)^{|C|\times\lambda}$.

4. Probabilistic Approximations Regions

The classical rough set theory is extended by variable precision rough set method (VPRS) in [7, 9, 10, 11], partial classification is taken into account by introducing an error probability threshold $\beta \in [0, 0.5)$, and it identifies all the condition classes with any decision class if the error ratio is not higher than this threshold. Given the approximation space (U, IND), an *absolute certainty gain (gabs)* is proposed to qualify the degree of the dependency from determinative class to conditional class:

DEFINITION 3 *If Y is definable in (U, IND), then the absolute certainty gain between sets X and Y is given by:*

$$gabs(X|Y) = \frac{\left|\sum_{E\subseteq Y} P(E)P(X|E) - P(X)\sum_{E\subseteq Y} P(E)\right|}{\sum_{E\subseteq Y} P(E)}, \tag{6}$$

where $P(E)$ and $P(X)$ are the probabilities of conditional class E and determinative class X, and $P(X|E)$ is the conditional probability of X. All these probabilities are estimated by the ratios of cardinalities of the sample data. The probability of the rule $s \rightarrow t$ is depicted by $gabs(r_{X|Y}) = gabs(X|Y)$, while X and Y are the corresponding classes of s and t. With the symmetric limits proposed, a precision control parameter denoted with β is utilized to define the positive and negative regions of X. All the rules with approximation threshold β satisfied can guarantee the corresponding associated level of classification quality of the both approximation regions, and the domination of the approximation regions can be elucidated with $POS_\beta(X, \neg X) = \cup\{E : gabs(X|E) \geq \beta - P(X)\}$. Let $X, Y \subseteq U$ be a non empty set of objects, the error ratio of X pertinent to Y denoted by $c(X, Y)$ is defined as:

$$c(X,Y) = \begin{cases} 1 - |X \cap Y|/|X|, & |X| > 0, \\ 0, & |X| = 0. \end{cases} \tag{7}$$

The operator $||$ denotes the cardinality in short. For given $\beta \in [0, 0.5)$, therefore, the $B-lower$ and $B-upper$ approximations of X with threshold β derived from $B \subseteq A$ are $R_B^{\beta}(X) = \{x \in U | c([x]_B, X) \leq \beta\}$; $R_{\beta}^{B}(X) = \{x \in U | c([x]_B, X) < 1-\beta\}$ respectively. Here, $[x]_B$ denotes the equivalence class including $x \in U$ with respect to B and it can be expressed as: $[x]_B = \{E \in U/B | x \in E\}$.

From Proposition 2, the associate level of quality of the neighborhood can be dominated by a granulation threshold λ with predefined covering certainty satisfied, and it can be denoted with $I_\tau(x) = \cup\{y : P(x\tau y) \geq (1-\lambda)^{|B|\times\lambda}\}$. For a given threshold pair (λ, β), the neighborhood system (U, IC) like equation (2) can be generated so that each $I(x)$ includes any object y with no less matching probability than $(1-\lambda)^{|B|\times\lambda}$. Therefore, the absolute certainty gain between any sets X and the universe U can be depicted by $P(X)\ P(X|I_\tau(x))$ and $P(I_\tau(x))$ similarly with equation (6), and the graduation threshold β can be utilized to guarantee the associated level of quality of neighborhood of the approximation regions. To depict the elements of all three cases in Proposition 2, let $\overset{*}{E}(x,y) = \{c_j \in C | c_j(x) = c_j(y) = *\}$ denotes the attributes subset where x and y are unavailable, $E^*(x,y) = \{c_j \in C | c_j(x) = * \wedge c_j(y) \neq *\}$ and $E_*(x,y) = \{c_j \in C | c_j(x) \neq * \wedge c_j(y) = *\}$ denote the similar meaning. Because similarity relation has a unilateral effect on both $P^{\models}$ and the joint power compared with tolerance relation as shown in Proposition 2, the two-layer domination has different forms for both the "missing value" and "absent value" semantics.

For the "missing value" semantics, the unavailable value is just lost. Following the discussion in Proposition 2, the error ratio is

$$c_T(x,y) = 1 - \frac{|\overset{*}{E}(x,y)| + |E^*(x,y)| + |E_*(x,y)|}{2|C|}. \tag{8}$$

Let T_B be tolerance relation, then $T_B^{\lambda}(x)$ denotes the tolerance class of x according to the threshold of the credible granulation $\lambda \in (0.5, 1]$, and it can be denoted with $T_B^{\lambda}(x) = \{y \in U | T_B(x,y) \wedge c_T(x,y) > \lambda\}$. For the tolerance relation with the granulation and graduation credible threshold pair (λ, β), the two-layer approximation regions can be induced with the predefined associated level of quality of both the granulation and graduation certainty satisfied.

DEFINITION 4 *The two-layer lower and upper approximations of $X \subseteq U$ are*

$$X_{BT}^{\lambda\beta} = \left\{x \in U \left| \frac{|T_B^{\lambda}(x)| \cap |X|}{|T_B^{\lambda}(x)|} \geq \beta \right.\right\};\ X_{\lambda\beta}^{BT} = \left\{x \in U \left| \frac{|T_B^{\lambda}(x)| \cap |X|}{|T_B^{\lambda}(x)|} > 1-\beta \right.\right\}.$$

For the "absent value" semantics, the unavailable value is not to be considered. $\overset{*}{E}$ does not affect the granulation credibility as discussed in Proposition 2, the error ratio is

$$c_S(x,y) = 1 - \frac{|E^*(x,y)| + |E_*(x,y)|}{2|C|}. \tag{9}$$

Let S_B be non-symmetric similarity relation, then $R_B^{\lambda}(x)$ denotes the similarity class and $R_{\lambda}^{B}(x)$ denotes converse similarity class of x with the granulation credibility threshold $\lambda \in (0.5, 1]$ satisfied are $R_B^{\lambda}(x) = \{y \in U | S_B(y,x) \wedge c_S(x,y) > \lambda$ and

$R_\lambda^B(x) = \{y \in U | S_B(x,y) \wedge c_S(x,y) > \lambda\}$ respectively. For the non-symmetric similarity relation and the credibility threshold pair (λ, β), the two-layer approximation regions can be induced as:

DEFINITION 5 *The two-layer lower and upper approximations of $X \subseteq U$ are*

$$X_{BS}^{\lambda\beta} = \left\{ x \in U \left| \frac{|R_\lambda^B(x)| \cap |X|}{|R_\lambda^B(x)|} \geq \beta \right. \right\}; \; X_{\lambda\beta}^{BS} = \left\{ x \in U \left| \frac{|R_B^\lambda(x)| \cap |X|}{|R_B^\lambda(x)|} > 1-\beta \right. \right\}$$

We suppose any decision class D_w in $U/\{d\} = \{D_1, D_2 ... D_r\}$ according to $D_w = \{x \in U | d(x) = w\}$ where $w = \{1, 2 ... r\}$, then the universal form of decision class can be derived from this form through a simple transformation, so a decision rule can be presented in the following form where $D_{wB}^{\lambda\beta}$ denotes the positive region of D_w under relation τ.

PROPOSITION 6 *For the approximation space based on a neighborhood relation τ, each $x \in D_{wB}^{\lambda\beta}$ can induce a decision rule as:*

$$\bigwedge_{c_j \in B} (c_j, c_j(x)) \xrightarrow[\lambda]{\beta} d(x) = w. \tag{10}$$

5. Comparisons and Experiments

Compare the model in Definition 4-5 with their counterparts in Chapter 3. Our approximation model will provide some advantages. Let $BN_B^T(X) = X_T^B \backslash X_B^T$, $BN_B^S(X) = X_S^B \backslash X_B^S$, $BN_{BT}^{\lambda\beta}(X) = X_{\lambda\beta}^{BT} \backslash X_{BT}^{\lambda\beta}$ and $BN_{ST}^{\lambda\beta}(X) = X_{\lambda\beta}^{ST} \backslash X_{ST}^{\lambda\beta}$, we give the most important properties as follows:

(1) $X_{BT}^{0,1} = X_B^T$; $X_{0,1}^{BT} = X_T^B$; $X_{BS}^{0,1} = X_B^S$; $X_{0,1}^{BS} = X_S^B$

(2) $X_B^T \subseteq X_{BT}^{\lambda\beta} \subseteq X_{BS}^{\lambda\beta} \subseteq X \subseteq X_{\lambda\beta}^{BS} \subseteq X_{\lambda\beta}^{BT} \subseteq X_T^B$

(3) $X_B^S \subseteq X_{BS}^{\lambda\beta} \subseteq X \subseteq X_{\lambda\beta}^{BS} \subseteq X_S^B$

(4) $BN_{BT}^{\lambda\beta}(X) \leq BN_B^T(X)$; $BN_{BS}^{\lambda\beta}(X) \leq BN_B^S(X)$

From the definition of the two-layer probabilistic model in the former two sections, property (1) is obvious.

For property (2), the partial order of the lower approximation based on tolerance between the two-layer and classic method(denoted by $X_B^T \subseteq X_{BT}^{\lambda\beta}$) can be considered into the two threshold respectively. Because $\lambda \in [0.5, 1)$, some objects which partially match x are in $I_B(x)$ while not in $T_B^\lambda(x)$. All objects in $I_B(x)$ must belong to $T_B^\lambda(x)$ because of the complete inclusion of the objects in tolerance relation class, and it results in $T_B^\lambda(x) \subseteq I_B(x)$. For each $x \in U$, suppose $I_B(x) \subseteq X$, then $T_B^\lambda(x) \subseteq X$ and the reverse does not exist because the objects in $T_B^\lambda(x) \backslash I_B(x)$ do not definitely belong to X. From the analysis above, supposing $X_{BT}^\lambda = \{x \in U | T_B^\lambda(x) \subseteq X\}$, we assert $X_B^T \subseteq X_{BT}^\lambda$. Given the neighborhood noted by $T_B^\lambda(x)$ of any object in the universe, $X_{BT}^{\lambda\beta}$ includes all the objects according to $|T_B^\lambda(x) \cap X| / |T_B^\lambda(x)| \geq \beta$

but X_{BT}^{λ} only includes the objects whose neighborhood outright belong to X, so the existence of the objects supports $1 > |T_B^{\lambda}(x) \cap X|/|T_B^{\lambda}(x)| \geq \beta$, and it results in $X_{BT}^{\lambda} \subseteq X_{BT}^{\lambda\beta}$. From all the above, we assert $X_B^T \subseteq X_{BT}^{\lambda\beta} \subseteq X$. The proof of $X_{BT}^{\lambda\beta} \subseteq X_{BS}^{\lambda\beta}$ is similar with $X_B^T \subseteq X_B^S$ in [6], so we assert $X_B^T \subseteq X_{BT}^{\lambda\beta} \subseteq X_{BS}^{\lambda\beta} \subseteq X$; also because the upper approximation is symmetric to the lower approximation, property (2) is proven.

Similar with the above proof, the partial order relation between $R_{\lambda}^{B}(x)$ and $I_B(x)$ can induce property (3).

We can deduce property (4) from property (2) and (3).

Example in Table 1 is introduced to compare the tolerance, similarity and probabilistic approximation model. The probabilistic approximations based on similarity are a refinement of those obtained under the tolerance relation as shown in property (4), so the following analyzes the probabilistic method only based on tolerance relation and the results of the former is better.

Table 1. An given incomplete table

C	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}	a_{11}	a_{12}
c_1	3	2	2	*	*	2	3	*	3	1	*	3
c_2	2	3	3	2	2	3	*	0	2	*	2	2
c_3	1	2	2	*	*	2	*	0	1	*	*	1
c_4	0	0	0	1	1	1	3	*	3	*	*	*
d	Φ	Φ	Ψ	Φ	Ψ	Ψ	Φ	Ψ	Ψ	Φ	Ψ	Φ

For the above system $DS = \{U, B; \{d\}, V, f\}$, the results of two-layer tolerance neighborhood and approximation regions with the threshold pair $(0.6, 0.6)$ are:

$T_B^{\lambda}(a_1) : (a_1, a_{11}, a_{12})$, $T_B^{\lambda}(a_2) : (a_2, a_3)$, $T_B^{\lambda}(a_3) : (a_2, a_3)$,
$T_B^{\lambda}(a_4) : (a_4, a_{12})$, $T_B^{\lambda}(a_5) : (a_5, a_{12})$, $T_B^{\lambda}(a_6) : (a_6)$,
$T_B^{\lambda}(a_7) : (a_7, a_9, a_{12})$, $T_B^{\lambda}(a_8) : (a_8)$, $T_B^{\lambda}(a_9) : (a_7, a_9, a_{11}, a_{12})$,
$T_B^{\lambda}(a_{10}) : (a_{10})$, $T_B^{\lambda}(a_{11}) : (a_1, a_9, a_{11})$, $T_B^{\lambda}(a_{12}) : (a_1, a_4, a_5, a_7, a_9, a_{12})$;
$\Phi_{BT}^{\lambda\beta} = (a_1, a_4, a_7, a_{10}, a_{12})$; $\Phi_{\lambda\beta}^{BT} = (a_1, a_2, a_3, a_4, a_5, a_7, a_9, a_{10}, a_{12})$;
$\Psi_{BT}^{\lambda\beta} = (a_6, a_8, a_{11})$; $\Psi_{\lambda\beta}^{BT}|(0.6, 0.6) = (a_2, a_3, a_5, a_6, a_8, a_9, a_{11})$.

Table 2. The cardinality of approximation result

App	T_B	S_B	*Probabilistic method with pair*(λ, β)			
			(0.6,0.6)	*(0.6,0.7)*	*(0.7,0.6)*	*(0.7,0.7)*
$\|L(\Phi)\|$	*0*	*2*	*5*	*3*	*5*	*3*
$\|H(\Phi)\|$	*11*	*9*	*9*	*10*	*8*	*8*
$\|L(\Psi)\|$	*1*	*3*	*3*	*2*	*4*	*4*
$\|H(\Psi)\|$	*12*	*10*	*7*	*10*	*7*	*9*

Given different threshold pairs, the cardinalities of the lower and upper approximations through two-layer method are listed in Table 2 for comparison, where T_B and S_B are classical methods. Except for the properties mentioned above, relative to the existent methods, the probabilistic method has another two advantages to control the boundary concerning the precision thresholds quantitatively. The probability based two-layer method can increase the positive region and decrease the boundary through the threshold pair (λ, β); the quality of the two-layer method decreases while λ rises and the quality rises while β rises.

6. Conclusion

This paper begins with the functions of approximation sets in the view of decision reducing, and then it expands the boundary approximation with a probabilistic method though two layers. The major contribution of this paper is to control the cardinality of the rough set of decision class under both the granulation credibility threshold and the graduation credibility threshold. More topics such as reducing algorithm development and the experimental analysis of parameters on large real-life dataset will be done in further research.

References

1. Pawlak Z. (1991) Rough Sets: Theoretical Aspects of Reasoning about Data. Kluwer Academic Publishers, Dordrecht, Vol. 9.
2. Stenfanowski J., A. Tsoukias, (2001) Incomplete information tables and rough classification. *Computational Intelligence*, 17:454-466.
3. Skowron A., Slowinski R., Synak P. (2005) Approximation spaces and information granulation. *T. Rough Sets*. pages 175-189
4. Kryszkiewicz M. (1998) Rough set approach to incomplete information system. *Information Sciences*. 112:39-49.
5. Slowinski R., Vanderpooten D. (2000) A generalized definition of rough approximation based on similarity. *IEEE Transactions on Data and Knowledge Engineering*.
6. Stefanowski J., Tsoukias A. (1999) On the extension of rough sets under incomplete information. In: *N. Zhong, A. Skowron, S. Ohsuga, (ads.), New Directions in Rough Sets, Data Mining and Granular Soft Computing*. LNAI 1711:73-81.
7. Ziarko W. (2005) Probabilistic Rough Sets. *RSFDGrC*. 1:283-293.
8. Wang G. Y. (2002) Extension of Rough Set Under Incomplete Information Systems. *Journal of Computer Research and Development*. 39(10):1238-1243.
9. Gong Z. T., Sun B. Z., Shao Y. B., Chen D. G., He Q. (2004) Variable Precision Rough Set Model Based on General Relation. *Proceedings of the Third International Conference on Machine Learning and Cybernetics, Shanghai*. pages 26-29.
10. Wang J. Y., Zhou G. C. (2005) Variable Precision Rough Set model in Incomplete Information System. *Proceeding of the Fourth International Conference on Machine Learning and Cybernetics, Guangzhou*, pages 1883-1887.
11. Mi J. S., Wu W. Z., Zhang W. X. (2004) Approaches to knowledge reduction based on variable precision rough set model. *Information Sciences*. 159:255-272.
12. Lenarcik A., Piasta Z. (1887) Probabilistic rough classifiers with mixture of discrete and continuous attributes. In *Lin T.Y., Cerone N.(ed.), Rough sets and Data Mining, Kluwer Academic Publisher*. pages 373-390.